Modelling the Cocktail Party: A Binaural Model for Speech Intelligibility in Noise

Sam M. Jelfs

Thesis submitted to Cardiff University
for the degree of Doctor of Philosophy

January 2011

Declaration

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ..... (candidate) Date.....

Statement 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed ..... (candidate) Date.....

Statement 2

This thesis is the result of my own independent work / investigation, except where otherwise stated. Other sources are acknowledged by explicit reference.

Signed ..... (candidate) Date.....

Statement 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to other outside organisations.

Signed ..... (candidate) Date.....

Statement 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access previously approved by the Graduate Development Committee.**

Signed ..... (candidate) Date.....

Contents

List of Figures

List of Tables

List of Symbols

AI	Articulation Index, page 4
AM	Amplitude Modulation, page 6
B_{BE}	Benefit due to Better-Ear Listening, page 53
B_{BU}	Benefit due to Binaural Unmasking, page 53
BRIR	Binaural Room Impulse Response, page 22
CI	Cochlear Implant, page 112
CVC	Consonant-Vowel-Consonant, page 17
D/R	Direct-to-reverberant Ratio, page 4
E-C	Equalization-Cancellation Theory, page 9
F_0	Fundamental Frequency, page 155
HRTF	Head Related Transfer Function, page 33
ILD	Interaural Level Difference, page 3
IM	Informational Masking, page 3
IPD	Interaural Phase Difference, page 8

IR	Impulse Response, page 52
ITD	Interaural Time Differences., page 2
KEMAR	Knowles Electronics Manikin for Acoustics Research, page 33
MTF	Modulation Transfer Function, page 17
r	Correlation Coefficient, page 16
RMSE	Root Mean Square Error, page 31
RT_{60}	Reverberation Time, (60 dB), page 4
s.d.	Standard Deviation, page 17
SII	Speech Intelligibility Index, page 3
SRT	Speech Reception Threshold, page 7
STI	Speech Transmission Index, page 3
TIR	Target-to-Interferer Ratio, page 24
TIR_{BE}	Target-to-Interferer Ratio with Better-Ear listening., page 52
TIR_{BE+BU}	Target-to-Interferer Ratio with Better-Ear Listening and Binaural Unmasking., page 53
TIR_{OMNI}	Omnidirectional Target-to-Interferer Ratio., page 52
U_{80}	Useful-to-Detrimental ratio, with a useful time of 80 ms., page 165

Acknowledgements

Firstly, I wish to thank Dr Mike Fedeski from the Welsh School of Architecture and Prof John Culling from the School of Psychology. Between them they have made my time studying at Cardiff University a pleasure, stimulating my interest in psychoacoustics, guiding me in my research, and challenging me as I have gone forward. As neither an architect nor a psychologist I feel privileged to have been allowed the chance to take on this project, one that was some way outside the field of my previous studies.

I also wish to thank Dr Mathieu Lavandier and Dr Mickael Deroche for their stimulating discussions, both in the weekly lab meetings and, along with Joel Burton, in the more relaxed surroundings of the local pub.

Finally, I wish to thank my family, without their support I would never have got as far as I have. In particular Beth, for finally submitting when she did, it would have been embarrassing to have submitted before her. I also thank Rae, for putting up with me this past year, through all the stress and late nights of writing up without comment or complaint.

Summary

We often listen to speech in an imperfect environment, with noise and reverberation; there will be voices around us, in complex acoustics. In this "cocktail-party" situation (Cherry 1953) listeners are helped by two binaural processes to segregate the desired voice from the competing noise: Better-Ear listening (BE) and Binaural Unmasking (BU). The aim of this thesis was to develop a model capable of efficiently predicting the benefits of BE and BU from Binaural Room Impulse Responses (BRIR). The developed model is a computationally efficient version of that created by Lavandier & Culling (2010) that predicts speech reception thresholds which include the benefit of binaural-unmasking, as explained by Equalization-Cancellation theory (Durlach 1963, 1972), and the benefit of better-ear listening, through Target-to-Interferer ratio analysis. The model accurately predicted a number of appropriate data sets from the literature that measure speech reception thresholds as a function of target and interferer source locations. Application of the model to a number of novel situations allowed environmental factors affecting intelligibility to be predicted and explored.

In most situations, the effect of reverberation is to reduce the level of BE and BU, except when the listener is close to the interfering source, but this is when the benefits are needed the most. Depending on the spatial separation and source distances, the inclusion of multiple interferers again reduces the benefits in the majority of situations. To examine the benefits of head orientation a number of configurations were tested, whilst rotating the listener relative to the sound field. Benefits exceeding 12 dB can be achieved through modest rotations, particularly showing the benefits of BE. According to the model, the current literature on the benefits of bilateral cochlear implantation has underestimated that benefit by employing sub-optimal spatial configurations; using optimum orientations the model predicts an extra 6 dB of benefit being available to the listener. In a simulated restaurant situation, the model predicts that orientation of a table can affect the ability of a listener by up to 5 dB.

Chapter 1

Introduction

Often, when we listen to a person talking, it is not done in a "perfect" listening environment, one that is devoid of other noise and excessive reverberation. Rather, it is more likely that we have a number of competing voices from other people around us talking, as well as other sources of noise, such as air conditioning, whilst this is all taking place in a reverberant space. Yet even with all the competing and interfering noise within the room, it is often possible for two people to hold a conversation. This situation is commonly referred to as the "cocktail-party problem" (Cherry 1953). The focus of the present work is to develop a method of predicting this ability of a listener to understand speech in such a situation, and to be able to efficiently analyse spaces to see if there are ways in which architects and acousticians can enable the listener to do this more comfortably.

Cherry (1953) identified a number of discriminating cues that allow for the separation of competing voices, principally: visual cues; speech characteristics; translational probabilities; and spatial separation. Visual cues are things such as lip reading and hand gestures that facilitate the ability to communicate, whilst speech characteristics are differences in items such as gender, pitch, speed, and accents. Translational probabilities

are cues such as the topics of conversation which allow the listener to better understand speech based on the predictability of the content. Spatial (azimuthal) separations are the focus of the present work.

An azimuthal (horizontal) separation of sound sources has been shown to allow the listener to better understand speech in a background of noise (Plomp 1976, Bronkhorst & Plomp 1988, Bronkhorst & Plomp 1992, Peissig & Kollmeier 1997, Culling, Hodder & Toh 2003, Hawley, Litovsky & Culling 2004), this is a phenomenon known as "spatial unmasking". The ability to localize a sound source, and therefore the ability to use spatial unmasking to discriminate between two or more sources, is assisted by two cues: interaural level differences and interaural time differences (Bronkhorst & Plomp 1988).

Interaural Time Differences (ITDs) are caused by the difference in path lengths between the source and the two ears. For a source in front of the listener the distance to the two ears will be equal, resulting in no ITD. For a source directly to one side of the listener the path to the ear on the opposite side as the source will be longer than that of the ear on the same side. At low frequencies, where the wavelength of the sound is long compared to the size of the listener's head, the sound passes the head as if it was not there, whilst at high frequencies, where the wavelength is shorter than the size of the head, the sound travels around the circumference of the head. Differences in the ITDs of the signals received from the target source and any interfering noise sources can facilitate the perceptual separation of the target from the interferer through a process called "binaural unmasking" (Hirsh 1948, Licklider 1948). Binaural unmasking makes use of differences in target and interferer ITDs to perceptually reduce the sound level of the interfering source. Binaural unmasking, and a model to predict the benefits due to binaural unmasking, are examined in section 1.2.1.

Interaural Level Differences (ILDs) are mainly caused by the simple effect of the path between the sound source and one of the ears being occluded by the listeners head, causing a "head-shadow" effect. This potentially creates a beneficial target-to-interferer ratio at the shadowed ear, dependent on the location of the target and the interfering sources.

It should be noted that when we discuss speech intelligibility with relation to interferers, we need to be careful to specify the type of interferer present. With a continuous noise interferer, such as an air handling unit, there is typically only "Energetic Masking" (EM) present, this being the effect of the energy of the interferer masking the energy contained within the target. When the interferer is speech there is also the effect of "Informational Masking" (IM) (Kidd, Mason, Brughera & Hartmann 2005, Shinn-Cunningham, Ihlefeld, Satyavarta & Larson 2005). Informational masking has been something of an "umbrella term" but generally speaking deals with difficulties attending to a single voice due to the similarities in the target and interferer, or the uncertainty over which speech is the target and which the interferer. (See Watson (2005) for a discussion on the use of the term "Informational Masking".) The present work considers only the effects of energetic masking and makes little reference to informational masking. This is due to the sole use of noise interferers, which typically do not produce IM effects.

1.1 Current Prediction Methods

Presently there are two main methods of predicting the intelligibility of speech in noise or reverberation, either from real-world measurements or computer predictions, that are employed by architects and acousticians when designing spaces or sound reinforcement systems. They are the Speech Intelligibility Index (SII), and the Speech Transmission Index (STI). Other measures of acoustic quality are available, such as the Direct-

to-Reverberant Ratio (D/R), and the Reverberation Time (RT_{60}). Whilst they both give information regarding the quality of the acoustics within the space, they do not directly relate to the intelligibility of speech, especially when an interfering source is present.

For both the SII and the STI, noise is considered to be a constant steady-state source. For a non-continuous source, or one that changes over time, its level and effects are averaged over time prior to calculation.

1.1.1 Speech Intelligibility Index

The Speech Intelligibility Index is set out in the American National Standard, "S3.5 - Methods for calculations of the Speech Intelligibility Index" (ANSI 1997), which is a revision of the earlier standard, "Methods for the Calculation of the Articulation Index" (ANSI 1969).

The earlier Articulation Index (AI) is based upon the work of Harvey Fletcher and colleagues at AT&T's Western Electric Research Lab (which would later became Bell Labs). The research carried out at AT&T was focussed on the prediction and measurement of the effects of changes to telephone circuitry on the perception of speech understanding (Fletcher & Galt 1950). The work of Fletcher and his colleagues was summarised by French and Steinberg (1947), which became the basis of the Articulation Index standard (Kryter 1962*a*, Kryter 1962*b*, ANSI 1969).

To calculate the Articulation Index one must first independently measure the levels of the target speech and the noise interferers within twenty prescribed frequency bands, and then sum the signal-to-noise ratios within each band, dividing the overall result by a given, frequency dependent, constant. The result of this process is a figure, ranging from 0 to 1, that relates to the percentage of speech cues available to the listener. It must be made

clear that a score of 0.5 is equal to having 50% of the speech cues available to the listener, not that the listener will understand 50% of what is being said. In fact, with 50% of the speech cues a normal hearing listener will likely understand almost everything that is being said (Kryter 1962*b*). To relate the AI figure to the percentage of speech understood it is necessary to have a function that can convert from one score to the other.

Given that the Articulation Index was developed in the telephone industry it is understandable that it does not take into account any of the factors that affect binaural hearing, but this clearly makes it unsuitable for predicting the ability of a listener when they do have the ability to use both of their ears. Studies have shown large differences in listener ability between using binaural and monaural hearing (e.g. Hawley *et al* (2004)).

The revised Articulation Index, the Speech Intelligibility Index, uses essentially the same principle as that used by the AI, though with a greater flexibility over the bands used, and inclusion of effects such as vocal effort, as well as the free-field to eardrum transfer function, which is included to translate between what is measured in the free field, and what would be received at the eardrum of the listener. There are a number of conditions that have to be met for the SII to be accurate:

1. *Listener faces both the speech and noise source in an otherwise free field* [i.e. no other sources present]*or noise and speech are assumed to be omnidirectional sound sources* [i.e. arriving at the listener from all directions]
2. *Listening is monaural*
3. *The speech and noise sources are independent of each other and their properties can be accurately measured in the absence of each other."*

(ANSI 1997, p 4. §3.5)

For real-world situations conditions 1 and 2 will rarely be met, even approximately. If the listener is facing both the target and the interferer they must be in the same direction, a coincidence which would only rarely occur. Secondly, apart from telephone conversations, it is unlikely that a normal hearing person will be listening to a target and an interfering source monaurally. There are two notes within the ANSI (1997) document (notes 3.11.1 and 3.15.1) which say that, for binaural listening conditions, the equivalent speech spectrum level and noise spectrum level, factors which are both used as part of the calculation, may include appropriate corrections for the "effects of binaural asymmetry", but there is no suggestion as to what these corrections may be.

1.1.2 Speech Transmission Index

The Speech Transmission Index (BSI 2003, Steeneken & Houtgast 1980) takes a different approach from that of the Speech Intelligibility Index. Instead of measuring the physical level differences between the target and the interfering noise sources it uses a special modulated test signal. It is based on the principal that the intelligibility of speech transmitted through a "system", be it a telephone network, or two people talking within a room, is correlated with the preservation of the Amplitude Modulation (AM) in each frequency band. AM is the way in which the amplitude (level) of a signal varies with time. By measuring the levels of modulation in the test signal prior to transmission and measuring the levels of modulation again upon reception the degradation of the signal can be calculated, which can be used to predict the corresponding degradation in intelligibility.

The STI was designed to predict the intelligibility of speech in diotic listening (that is, where the same signal is received at both ears, though this is more or less equivalent to the monaural listening assumed by the SII / AI, where the signal is received at one

ear only). However the benefits of binaural listening, which requires dichotic stimulation (different signals at each ear) are still not represented. If the target and interfering noise sources are located at the same position directly in front of the listener, as is assumed in both the STI and the SII standards, then the differences between monaural, diotic and dichotic listening will be minimal. The predicted intelligibility will be accurate, but only for these limited situations. If the sources are separate and located other than in front of the listener, the most likely real-world situation, then the differences between monaural, diotic and dichotic listening may be high, and the STI will be inaccurate.

Both the SII and the STI give a score between 0 and 1 as their output, which does not directly link to speech intelligibility, but rather needs to be passed through a situation-specific transfer function to allow for direct comparison to measured speech intelligibility scores, such as the percentage correctly scored by participants listening to sentences in noise. It is desirable to be able to create a method of predicting intelligibility that can be directly linked to measured results, such as the Speech Reception Threshold (SRT), which gives a target-to-interferer ratio in decibels required for the listener to identify 50% of words in spoken sentences correctly. The benefit of this is to enable the user of such a method to easily convert between the result of the method and the level of intelligibility, without the need for functions that can convert from one measure to another.

In practice both the SII and the STI can be estimated either from 3D computer models of spaces using acoustic analysis software, or from real-world measurements.

1.2 Binaural Hearing

The fact that humans have two ears allows us to not only receive different signals at each ear, but also to process those two signals independently and in conjunction with

each other. One of the benefits of binaural hearing is sound localisation, the ability of the listener to determine the location of the source of a sound. There are two key ways in which binaural hearing contributes to this process, through using either Interaural Time Differences or Interaural Level Differences. ITDs can also be expressed as Interaural Phase Differences (IPDs), where, due to the extra time differences taken for the sound wave to arrive at the more distant ear, it does so at a different point in its wave cycle, and as such has a different phase at each ear.

For the understanding of speech in a background of noise, the ability to localise sound, and the existence of interaural time / phase, and level differences allows the listener to deploy two beneficial processes: binaural unmasking and better-ear listening.

1.2.1 Binaural Unmasking

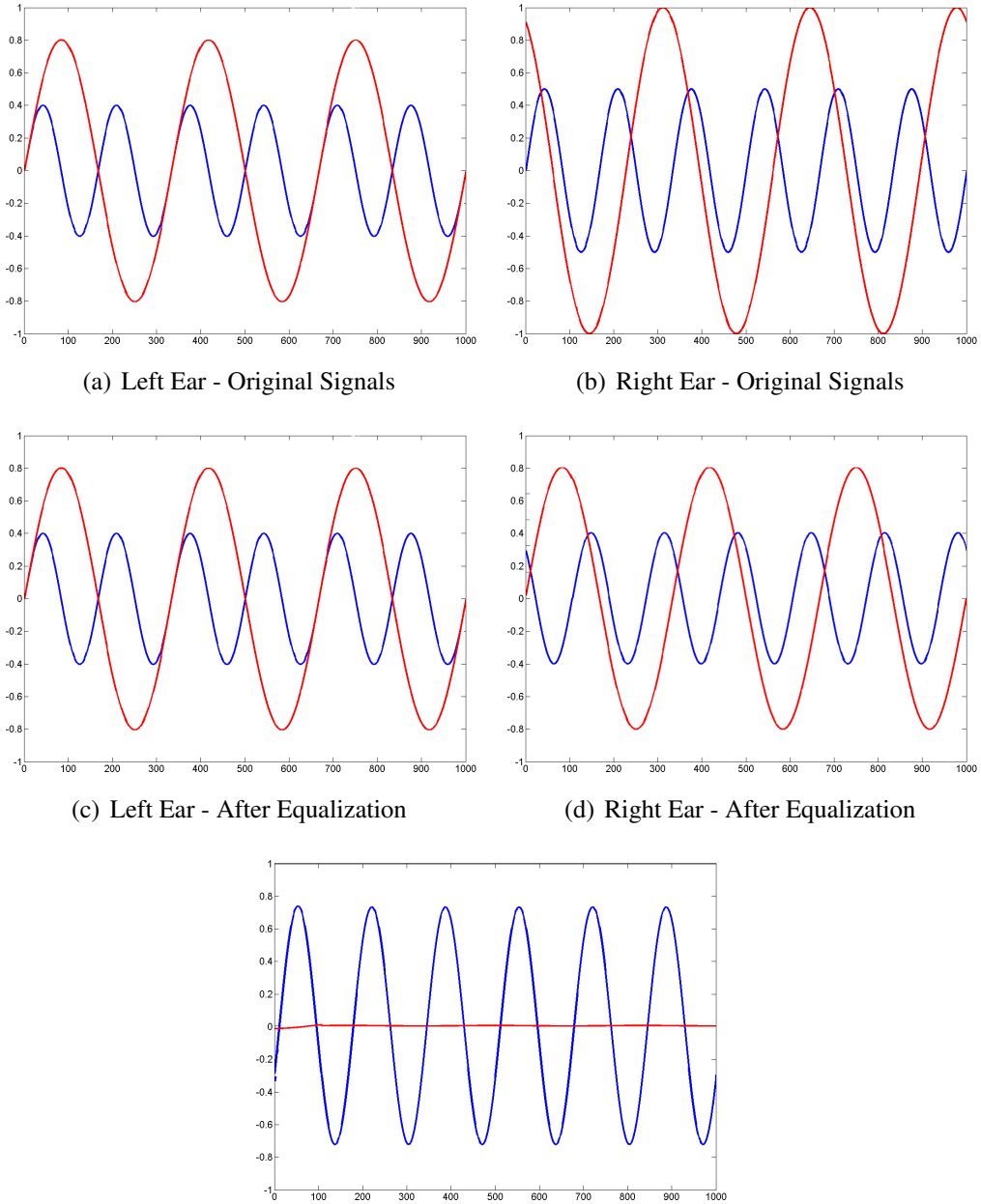
> *"If a communication engineer, confronted with a sound wave consisting of speech mixed with audible random noise, were requested to build a device to separate the speech from the noise, he would be hard pressed to produce a mechanism as effective as the human auditory system."* (Licklider 1948)

A number of studies have looked at the ability of human listeners to separate target sounds from competing noise, either monaurally, diotically, or binaurally. For listening binaurally, the participants perform better when there is a difference in IPDs between the target signal and the interfering noise, than when they have the same IPD. This has been observed using synthesised differences to detect tones in noise (Hirsh 1948), to understand speech in noise (Licklider 1948) and by using naturally occurring ITDs in the sound source due to its location to understand speech in noise (Bronkhorst & Plomp 1988). This ability to process time / phase differences when listening binaurally to better understand

or detect a target signal is known as binaural unmasking. One theory to explain the way in which the human brain is able to do binaural unmasking is through the Equalization-Cancellation (E-C) theory (Durlach 1963, Durlach 1972).

Equalization-Cancellation theory is a "black box" model to predict the Binaural Masking Level Difference (BMLD). The BMLD is the difference in listeners' detection threshold between when they are listening to a configuration with no difference in IPD between the target and the interferer, and when there is a given difference in IPD, either caused due to the sources having different spatial orientations or through modifications of the test material.

E-C theory works on the premise that the brain is able to take the two signals arriving from the two ears and delay the combined target and interferer signal in one of the ears so that the interferer in that ear is in phase with the interferer in the opposite ear. Figure 1.1 shows a graphical representation of this using sine waves. Panels (a) and (b) show the signals at the left and right ear respectively, with the target signal in blue, and the interfering signal in red. The target signals arrive at the two ears at the same time giving identical traces on the plots, whilst the right interferer arrives 1/4 of a cycle before the left, producing the different traces. Panels (c) and (d) show the signals after both the target and the interferer in the right ear have been delayed so that the left and right ear interferers are the same. The right ear signals have also been reduced in level, so as to match the level of the signals in the left ear. This is the Equalization part of the model. The theory then subtracts the combined signal at one ear from the combined signal at the other ear, this being the Cancellation part of the model. If the interferer at the two ears has the same phase after equalisation, it will be cancelled. If this is done with total precision, the interfering signal will be entirely eliminated, as shown in panel (e), but the process is



(a) Left Ear - Original Signals

(b) Right Ear - Original Signals

(c) Left Ear - After Equalization

(d) Right Ear - After Equalization

(e) Output after Cancellation

Figure 1.1: Diagram of E-C Theory using simple sine waves. The horizontal axis is time in samples, and the vertical axis is amplitude. Blue traces are target signals, red traces are interferers. Original right ear interferer leads the left by 1/4 of a cycle. Equalization shifts both the right ear target and interferer so left and right interferers become in line. Cancellation subtracts the right ear from the left, cancelling the interferer, and preserving the target.

assumed to be inherently noisy, due to internal limitations on the estimation of the time and level differences in the signal, meaning that some of the interferer would remain. How well the target remains is dependent on the difference in interaural phase of the target after equalisation. In the graphical example of fig 1.1 the target is $180°$, or 1/2 a cycle, out of phase after equalization, and therefore it remains. If there is a smaller interaural phase in the target after equalization, cancellation will tend to remove the target and the interferer together, and with no target interaural phase difference after equalization both will be cancelled.

For any given azimuthal angle[1] of a source to the listener a specific ITD will occur. This ITD is roughly independent of frequency, but for E-C theory to work it is a difference in phase that is required. Given that the phase of a sound is related to the wavelength of that sound, and as such the frequency, the IPD will differ strongly between frequency channels. As the IPD changes, the amount by which E-C theory will be able to cancel the interfering sound source will change, meaning that for any given ITD the benefit due to E-C will change across frequency.

For E-C theory to work, the interfering sound not only has to have a different interaural phase from the target speech, but it has to have a high level of interaural coherence. Interaural coherence is a measure of how correlated (similar) the signals received at the left and right ears are. With a coherence value of one, the signals will be identical to each other, and with a value of zero they will have no similarity at all. With a high level of coherence a large proportion of the interfering noise can be cancelled. A perfect system will be able to cancel all of an interfering signal if it has a coherence of one. The lower the level of coherence, the less of the interferer can be cancelled. Coherence is time

[1]Throughout this work azimuths are measured with $0°$ as directly in front of the listener, and increasing to the right of the listener, clockwise if viewed from above. Negative values are not used.

independent, the signals can reach the two ears at different times, but so long as the same signal reaches the two ears, it will have a high level of coherence. Mathematically the time difference is almost unlimited, but for a human head the maximum difference for a sound to reach both ears is approximately 650 μs, dependent on the size of the head. In an anechoic room, with a source directly to one side of the listener, there will be the maximum interaural time delay between the two ears, and there will also be a high level of coherence.

Binaural unmasking is frequency limited. For detecting tones with a large interaural phase difference in the presence of broad-band noise interferers, the ability of listeners decreases by approximately 13 dB between 500 Hz and 4 kHz (Hirsh & Burgeat 1958), that is, the level of tone has to be increased by approximately 13 dB to remain detectable. This reduction in detectability has been attributed to the decline in the ability of the ear to accurately detect the phase of a signal with increasing frequency (Durlach 1964, Zurek & Durlach 1987, Bernstein & Trahiotis 1996), without which the listeners are unable to make use of binaural unmasking.

Chapter 2 discusses the computational theory for coherence and of calculating the benefits attributed to binaural unmasking through E-C theory.

1.2.2 Head Shadow

The second key effect available to the listener for understanding speech in noise is the effect of head shadow on sound level. If both the target and the interferer are to the front of the listener the signals received at the two ears will be at the same level. If the interferer is moved to the side of the listener then the ear on that side of the head will receive a higher level of signal from the interferer than the ear on the opposite side of the head, as

the head is effectively "blocking" that ear from the source. In this situation, assuming the target and the interferer are at the same distance from the centre of the listener, and are of equal level, then the ear on the opposite side of the head to the interferer will receive a higher level of target than of interferer, giving a better target-to-interferer ratio at that ear. Through use of better-ear listening (Edmonds & Culling 2006) the listener is able to selectively choose which ear has the best target-to-interferer ratio, and attend to that one ear, benefiting from the effects of head shadow.

Aside from better-ear listening and binaural unmasking there are also spectral cues, where the sources have different frequency patterns in the left and right ears, predominantly caused by the shape of the pinna. They have been shown to play a less important role in the understanding of speech in noise (Mesgarani, Grant, Shamma & Duraiswami 2003).

1.3 Effects of Reverberation on Speech Understanding

Reverberation can be detrimental to the understanding of speech in two ways; by affecting the monaural intelligibility of the speech signal, and by disrupting the ability of the listener to make use of the benefits of binaural hearing.

For speech in a reverberant room, without any other interfering noise sources, the effect of reverberation is to temporally smear the speech signal. In normal speech, the amplitude of the signal varies greatly, with dips in the speech and hard onsets and offsets. Reverberation partially fills in these dips, and smears the onsets and offsets. If speech is seen as an amplitude-modulated signal, then this is comparable to applying a low-pass filter to the amplitude envelope. The amount of reduction in the amplitude-modulation of the signal is a good predictor of intelligibility (Steeneken & Houtgast 1980) and is the

basis of the Speech Transmission Index, as explained in section 1.1.2.

In reverberation, not only does the sound reach the listener by the direct path from the source, but it will also reach the listener through many reflections, taking many different paths, which will tend to randomise the interaural level and time differences. As the signals reaching the two ears will be composed of many different reflections arriving at the ears at various delays, the interaural coherence is likely to be decreased. In the presence of an interfering source, this will reduce the effectiveness of binaural unmasking as predicted by E-C theory. Plomp (1976) measured the sound level required for target speech to be just intelligible in the presence of interfering speech. In anechoic conditions the target could be up to 6 dB quieter when the sources were spatially separated than when they were collocated. In the presence of reverberation (RT_{60} of 2.3 s) this benefit was reduced to approximately 1 dB.

Whilst reverberation is generally detrimental to understanding speech, early reflections can be useful. The direct sound from a source reaches the listener first, but reflections arriving shortly after the direct sound can reinforce the direct sound without causing significant temporal smearing. For the target, this is beneficial, adding power to the signal, whilst it may be detrimental when occurring to the interferer as it will add power to the interferer, and will also decrease the coherence of it. Later reflections cause temporal smearing, which is detrimental to the intelligibility of the target. Measures exist that are able to take this into account, such as the Early-to-Late ratio and the Useful-to-Detrimental ratio. The latter measures the amount of energy occurring within a given period after the direct sound (normally 50 or 80 ms, termed C_{50} and C_{80} respectively, C standing for Clarity), and compares it against the amount of energy occurring after that period. Whilst these measures take into account the effect of early reflections, they do not

account for many of the other features that enable or hinder the understanding of speech in noise.

1.4 Architectural Implications

Generally either the Speech Transmission Index or Speech Intelligibility Index methods of evaluation are employed by architects, acousticians and public address system installers for designing and evaluating spaces for their suitability for speech understanding. Yet both the STI and the SII are limited in their accuracy and scope, and given that both prediction methods are based upon the early work of speech understanding in telephony they do not take into account the fact that humans have two ears, and binaural hearing.

Through producing a prediction method that is able to explore the intelligibility of speech in a background of noise interferers, whilst taking into account the effects of reverberation on the listener's ability to use the benefits of binaural hearing, it is hoped that spaces can be designed that can aid our ability to understand what is being said. This will be through better understanding the effects of room geometry, material usage and source placement on binaural hearing and speech understanding.

Whilst in general terms it could be said that reverberation is detrimental to speech understanding it is important not to take the view point that all reverberation needs to be removed. Depending on the exact situation under investigation, reverberation may be of use, through early echoes, or it may be the case that the speaker does not want a distant person to be able to listen to what is being said, but only a person standing close to them. Classically specific reverberation times have been quoted as the ideal times for a room, depending on it's purpose, be it for classical music (0.6 - 1 second for chamber music, 1.1 - 2.5 seconds for a symphony orchestra) or for speech (0.7 - 0.8 seconds, all times taken

from Everest (2001)). Rather than simply designing a room for a specific reverberation time it will be better to be able to accurately predict the effects of reverberation, and tailor the room to suit the needs of the user.

1.5 State of the Art

A number of models have been developed to account for the effects of binaural hearing as part of the SII and the STI, most notably the work of Beutelmann & Brand (2006) on the SII, and van Wijngaared & Drullman (2008) on the STI.

The work of Beutelmann & Brand (2006) is a preprocessing method that modifies recorded acoustic signals to take into account the factors affecting spatial hearing, prior to calculation of the SII. They include two uncorrelated internal masking noises to account for the individual hearing thresholds of the two ears, and also include a stochastic implementation of the equalisation-cancellation mechanism using Monte-Carlo simulations. Whilst they attain high levels of accuracy (the overall correlation coefficient (r) of the predicted and observed data in their published study is 0.95) their prediction method is computationally demanding. Given that they use actual acoustic signals, rather than statistical analysis, the amount of data to process is large. Moreover, the stochastic nature of modelling actual internal noise and the consequent need to average over many iterations makes the detailed analysis of large spaces a difficult and time consuming problem. A revised version of the model of Beutelmann & Brand (2006) has been proposed (Beutelmann, Brand & Kollmeier 2010) that removes the need for the Monte-Carlo simulations, as well reducing the need to resynthesise all of the input waveforms from the output of their gammatone filterbank, but this still takes acoustic waveform recordings as an input to the model, and also requires knowledge of the individual hearing threshold of

the person being modelled.

Van Wijngaarden & Drullman (2008) have developed an extension to the STI to incorporate the effects of binaural hearing. They recognised that whilst the need for incorporating the effects of binaural hearing in the STI is great, they also saw that there are a number of factors that need to be preserved from the original STI. These are mainly the need for relatively fast measurements (15 second test signals); accurate results in noise, reverberation, and nonlinear distortion (such as compressions added by a telecom system); the need for the model to be kept simple; and the need for it to be a feasible extension to current measurement devices.

The method of van Wijngaarden & Drullman calculates their binaural extension of the STI in 30 millisecond time frames across the length of the 15 second test signal, within three 1-octave wide frequency bands centred on 500, 1000, and 2000 Hz, which are the bands in which the most useful binaural interactions for speech understanding occur (Zurek 1993, Blauert 1996). Within each time frame, they calculate the interaural cross-correlation, and for each cross-correlation delay they compute the Modulation Transfer Function (MTF) for the cross-correlation value across time frames at each time delay. For each band they choose the highest MTF value, combining them to create a single overall STI value. Whilst using their binaural extension to the STI, van Wijngaarden & Drullman were able to predict a number of Consonant-Vowel-Consonant (CVC) word scores to a relatively high level of accuracy (reported as a Standard Deviation (s.d.) of 9.2% between their modelled data and a specified reference curve when using their full extension) but it adds at least 1500 extra calculations to the original calculations used in the STI (approximately 100, dependent on the method used) for every configuration. They also calculated a "better-ear STI" which, with a 3 dB correction factor applied,

could predict the same data to an almost equal level of accuracy (reported s.d. of 10.6%) whilst only doubling the number of calculations needed. When trying to develop a system that can quickly and accurately analyse large spaces, these extra calculations will add considerably to the computational processing required, whether it be adding an extra 100 calculations, or an extra 1500.

Whilst both the work of Beutelmann & Brand (2006), and that of van Wijngaarden & Drullman (2008) propose extensions to the SII and the STI that will incorporate the benefits due to binaural hearing, they are both computationally heavy. The work of Beutelmann, Brand, & Kollmeier (2010) reduces some of the computaional overheads, but sill requires individual acoustic waveforms for the target and interfering sources, as well as *a priori* knowledge of the hearing threshold of the listener. The aim of this work is to create a method that can evaluate large architectural spaces quickly, and accurately, without the need for high levels of computer processing power. Any extension to a speech intelligibility measure needs to be suitable for use by architects, acousticians, and engineers, as well as by academics and researchers.

1.6 Aims of the Research

The aim of this research project is to develop an accurate method for predicting the intelligibility of speech in noise, taking into account the effects of binaural hearing, principally head-shadow and binaural unmasking. Any method developed will need to be validated against empirical data.

Once a model has been developed it will be used to investigate the ways in which the position and number of interferers affects speech understanding, as well as the affect that the acoustics of the room can help or hinder the listener.

The final aspect of binaural hearing and speech intelligibility to be investigated is the way in which the orientation of the listener, rather than position within a space, affects their performance.

In this introductory chapter the present methods used by acousticians to predict and measure the intelligibility of speech have been examined, and the rationale for developing a new method outlined. The theory of the effect of reverberation on speech and the ways in which the binaural system can facilitate our understanding of speech have also be discussed.

The second chapter will look at the prediction method to be used throughout the present study, and its development from an earlier method proposed by Lavandier and Culling (2010).

Chapter three uses this revised method to investigate some simple theoretical room scenarios, and the ways in which the level of reverberation, and the number of competing sound sources affects the listener's ability to use head shadow effects and binaural interactions to understand the target speech.

Chapter four looks at the effects of head orientation on the understanding of speech, and whether or not the classic scenario of facing the person to whom you are listening may not be the best orientation to adopt.

The results of chapters three and four are carried in to chapter five, where their implications are discussed with relevance to audiology, and in particular to people with hearing aids and cochlear implants.

Chapter six creates a more realistic listening situation than used in the previous chapters, looking at the ways in which, through architecture and acoustics, a more beneficial listening environment for holding conversations in restaurants can be created.

Finally chapter seven is a general discussion and conclusion of the present thesis, including ways in which the work can be expanded further.

Chapter 2

The Revised Model

One of the aims of this work, as set out in the introduction, was to develop a model capable of predicting the intelligibility of speech in the presence of one or more interferers, whilst including the effects of better-ear listening and binaural unmasking. One model currently able to do such a prediction is the model of Lavandier & Culling (2010), which though accurate is not efficient enough to be used with large data sets, such as in modelling many listener positions within a room. A more efficient implementation of this model is described below.

2.1 Methodologies, Revisions, and Advantages

2.1.1 Overview of the Lavandier & Culling Model.

Lavandier & Culling (2010), proposed a method to predict Speech Reception Threshold (SRT) measurements for speech in noise, within a room. Their method is a two-path model, as shown in figure 2.1(a). The first pathway calculates the expected binaural advantage due to binaural unmasking (Hirsh 1948, Licklider 1948) using Equalization-

Cancellation theory to predict the Binaural Masking Level Difference (BMLD) (Durlach 1963, Durlach 1972). The second path predicts the benefit of better-ear listening due to head shadow as a Target-to-Interferer ratio. This process also accounts for the effects of "room colouration" due to reverberation, that is, the frequency dependent level differences in the BRIRs caused by the position of the sources within, and the geometry of the room. Combined, the two paths account for the two cues associated with spatial unmasking (Bronkhorst & Plomp 1988), which is an established mechanism for the separation of competing sounds (Hawley et al. 2004, Plomp 1976).

The model of Lavandier & Culling (2010) takes an input of continuous noise samples with the long term average spectrum of speech, which are convolved by the relevant Binaural Room Impulse Response (BRIR) recordings, to create reverberant speech-shaped noise sources for both the target and the interferer. These waveforms are then processed through the two paths of their model independently.

Firstly, to calculate the level of binaural unmasking from the target and interferer waveforms they are passed through a gammatone filterbank (Patterson, Nimmo-Smith, Holdsworth & Rice 1987, Patterson, Nimmo-Smith, Holdsworth & Rice 1988) with two filters per Equivalent Rectangular Bandwidth (ERB) (Moore & Glasberg 1983). From each filter output four sections of the waveform are extracted, each 320 ms long (starting at 0.5, 1, 1.5, and 2 seconds respectively).

For each extracted section of waveform the left and right channels are cross-correlated using |WAVE (Culling 1996), with a 100 ms exponentially tapering temporal window, and a range of delays between $\pm$5 ms (the longest delay possible between the two ears of a human head). From the cross-correlation function the interaural coherence of the interferer (ρ, which is the maximum of the cross-correlation function) and the interaural

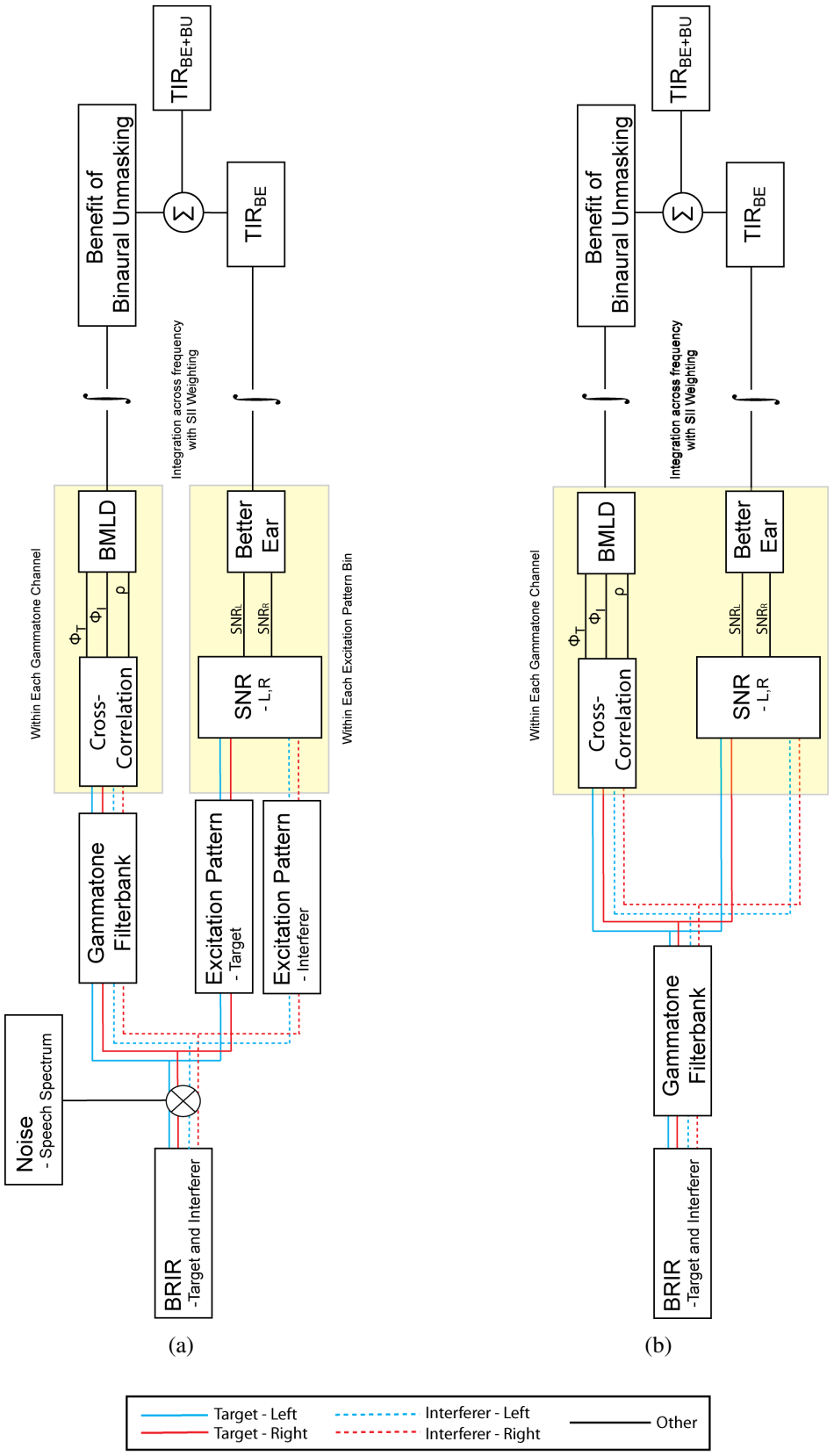


Figure 2.1: (a) The two-path model of Lavandier & Culling (2010), for predicting Speech Reception Thresholds, and (b) the proposed revised version.

phase differences of both target and interferer (Φ_T and Φ_I, which are the respective delays attributed to the maximum of the cross-correlation function multiplied by the filter centre frequency, given in radians) are calculated. From these three variables, ρ, Φ_T and Φ_I, along with the filter centre frequency, ω, (given in radians per second), the BMLD can be calculated using the method of Culling *et al* (2004, 2005) and the formula given in Culling *et al*, 2005:

$$BMLD = 10log(\frac{k - cos(\Phi_T - \Phi_I)}{k - \rho}) \tag{2.1}$$

where

$$k = (1 + \sigma_\epsilon^2)exp(\omega^2\sigma_\delta^2) \tag{2.2}$$

σ_δ^2 and σ_ϵ^2 are constants of 0.000105 and 0.25 respectively, and account for the processing errors for the level and time of the signals (from Durlach (1972)), without which perfect cancellation (and hence an infinite signal-to-noise ratio) would be possible.

Given the assumption that binaural thresholds are never below their equivalent monaural thresholds, (Durlach 1963), for any frequency channel where the formula returns a negative number the result is set to zero. To obtain the broadband binaural advantage the individual channel values are integrated across frequency using the weighting function from the Speech Intelligibility Index (SII) (ANSI 1997).

Secondly, to compute the benefits of better-ear listening in their model, Lavandier & Culling calculate the cochlear excitation patterns (Moore & Glasberg 1983) of each extracted section of the target and interferer waveforms, calculated between 0 and 33.25 ERBs (corresponding to 0 - 10kHz) every 0.13 ERBs for both the left and the right ears.

For each ear the Target-to-Interferer Ratio (TIR) is determined to be the difference between the target and interferer excitation patterns, with respect to frequency. The "better-ear" TIR at each frequency is taken to be the maximum of the left and right ear indepen-

dent TIRs, and the broadband target-to-interferer ratio as the integration across frequency of the maxima, again using the SII weightings.

For both the benefit of binaural unmasking and the better-ear target-to-interferer ratios, the values are averaged over the four sections of the waveform and combined to give one single "effective" target-to-interferer ratio, minimising the stochastic effects of convolving the BRIRs by random noise.

Some studies looking at binaural unmasking and better-ear listening have shown that the results obtained when participants are presented with the combined effects are smaller than the combination of the results when the effects are presented in isolation (Bronkhorst & Plomp 1988, Culling, Hawley & Litovsky 2004). Other studies have shown that the addition of the individual effects are equal to the combined effects (Zurek 1993, Hawley et al. 2004). Lavandier & Culling (2010) use the additive model to predict their collected data with a high level of accuracy, having a correlation coefficient of 0.97 between the modelled and the predicted values.

The result produced by the Lavandier & Culling (2010) model is a "benefit" of binaural hearing over listening through a single omnidirectional microphone. This can be considered to be an effective target-to-interferer ratio. For situations where there is a target and an interferer of equal power, in an anechoic space, and at an equal distance from the listener, a benefit of, for example, 3 dB may be predicted. This says that the listener has an effective target-to-interferer ratio of 3 dB. This is a relative measurement, and as such needs to be considered against a reference condition. The SRTs measured by Lavandier & Culling were done so using an adaptive threshold measure (Plomp & Mimpen 1979), they are therefore a target-to-interferer ratio by definition. An SRT of -3 dB says that the target can be 3 dB quieter than the interferer, yet still be 50% intelligible, this would cor-

respond with an effective target-to-interferer ratio of +3 dB. As such, a direct comparison can be made between the measured and the modelled data, requiring no calculation of indices (AI, SII) nor index-to-intelligibility mapping (Beutelmann & Brand 2006, Levitt & Rabiner 1967). To do this, the benefits, or effective target-to-interferer ratios, as predicted by the model need to be inverted, and a single value needs to be added to all of the modelled values, dependent on the reference condition of the measured data. In the Lavandier & Culling (2010) data this was the difference between the average level of the measured SRTs, and the average effective TIRs.

2.1.2 Revision of the Lavandier & Culling Model.

Whilst the model of Lavandier & Culling (2010) gave a high correlation with their data set it was desirable to make it computationally more efficient. The reason for this is principally a question of speed. When evaluating a small number of conditions, sixteen in the case of the Lavandier & Culling data set, differences in speed of the order of 1 second make little overall difference, yet when it comes to situations where there are many hundreds, if not thousands of sub-conditions within any one condition to be evaluated, such as when modelling listener positions across a regular grid within an acoustic environment, small increases in speed per calculation makes a big difference to the overall speed of the calculations. As the aim of the project was to produce a modelling methodology that could be used by architects and acousticians to evaluate the design of spaces and the effect that they have on the ability of persons within the space to hold a conversation it was essential that the model was as computationally efficient as possible.

To this end, two major changes have been introduced. Firstly the calculations take place directly on the binaural room impulse response recordings originally used to cre-

ate the target and interferer noise samples, and secondly the target-to-interferer ratio is calculated on the power of the signals at the output of the gammatone filterbank, rather than from cochlear excitation patterns. The generation of a cochlear excitation pattern requires the use of a second filterbank. Computationally it is more efficient to calculate the target-to-interferer ratio on the output of the gammatone filterbank already used to calculate the binaural unmasking component of the model than to include a second filterbank. The model has also been modified to allow for conditions containing multiple interfering sound sources.

The benefit of calculating directly from the binaural room impulse responses is that they can be much shorter than their respective convolved noise sources. Figure 2.2 shows a sample BRIR (upper panel) and a noise source convolved with the same BRIR. Both have a sampling frequency of 44.1kHz and have been passed through a gammatone filter with a 500 Hz centre frequency. The BRIR is 10,000 samples long, with a closely corresponding coherence of 0.4188, whilst the convolved noise is 1×10^6 samples long, with a coherence of 0.4183. However, if the length of the noise is reduced to 1×10^5 the coherence value is variable and only tends to match that of the BRIR to 1 decimal place. Due to the stochastic nature of the noise either multiple sections need to be evaluated, as in the method of Lavandier & Culling (2010), or long sections need to be used, both of which make the process much more computationally intensive compared to cross-correlating the BRIR directly. The only factor affecting the required length of the BRIRs is to ensure that they include sufficient amounts of the reverberant field; the more absorptive a space, the shorter the BRIRs need to be, whilst the noise would remain the same length. That said, with excessively short impulse responses there is a risk of the gammatone filters ringing, which may necessitate zero-padding of the IR. By evaluating the shorter BRIRs, as well

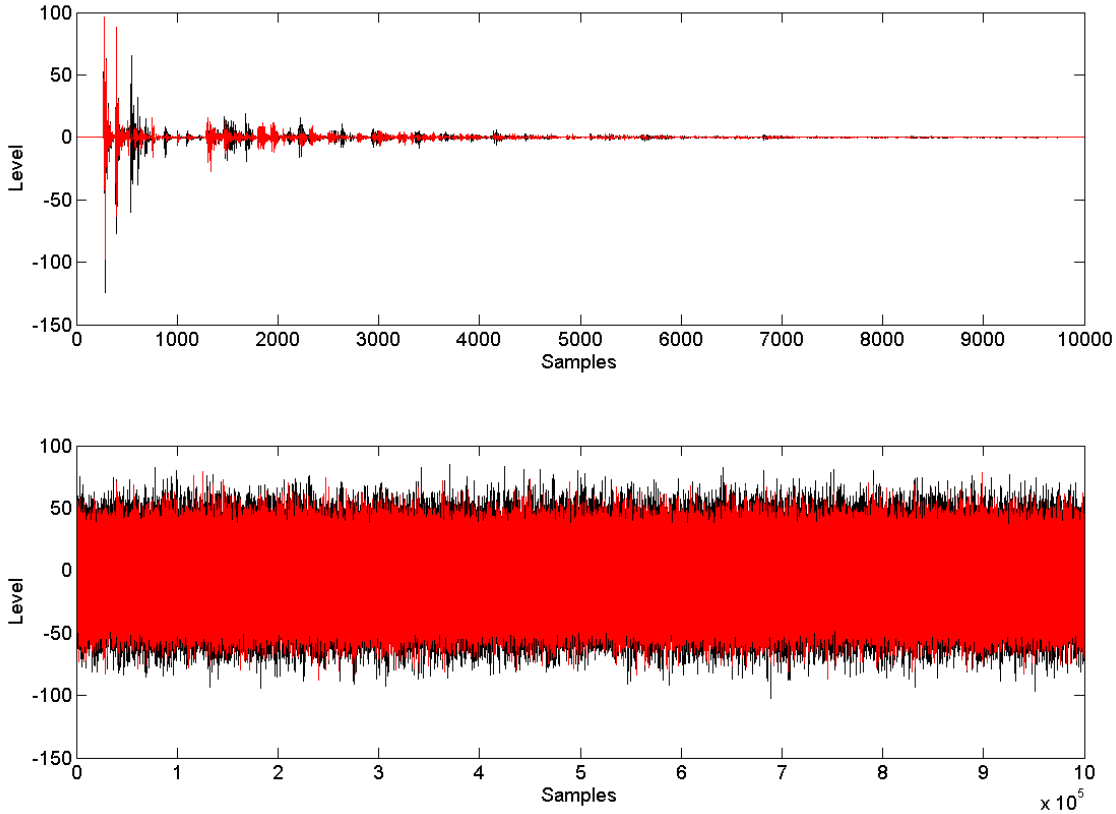


Figure 2.2: Comparison between a 500 Hz band filtered sample BRIR (upper panel) and 500 Hz band filtered noise convolved with the same BRIR (lower panel). Both have similar measured coherence values, but widely differing lengths. Left channel is shown in black, right channel in red.

as removing the need to carry out convolution of the speech shaped noise with the BRIRs, makes for a much more efficient process.

In their model Lavandier & Culling (2010) calculated the better-ear benefit due to room colouration by taking the difference of the target and interferer cochlear excitation patterns at the two ears. The excitation pattern is the distribution of internal excitation in the ear as a function of frequency, which "may be conceived at the output of each filter as a function of filter centre frequency."(Moore & Glasberg 1983, p. 752). Consequently, it is more efficient to take the output of the gammatone filterbank, since this has already been generated in the calculation of the level of binaural unmasking. Through calculation of the Root Mean Square (rms) level of the target and interferer within each filter channel the target-to-interferer ratio can easily be extracted. By removing the need to calculate the cochlear excitation patterns, and using only the gammatone filterbank, it is thus possible to further increase the speed of the calculations.

It has also been necessary to include a method of handling multiple interfering sound sources in the model, to allow for analysis of more complex listening environments. Fundamentally, the way in which the better-ear and binaural unmasking benefits are calculated does not change, but the set of interferer binaural impulse responses must first be combined into a single binaural pair. This is done by first concatenating the individual impulse responses together. The reason for concatenation rather than addition is to avoid waveform interaction between the different impulse responses. If addition was to be used the multiple signal samples that are in phase will increase the amplitude of that sample, whilst those that are out of phase would reduce the amplitude, giving undue bias to those samples that occur in phase. Another way to view the concatenation is that the normalised crosscorrelation of the concatenated signal gives the same result as adding

the non-normalised crosscorrelations of the individual binaural pairs, then averaging and normalising the resultant. Concatenation and crosscorrelation (with suitable limits to the range of lags analysed) is computationally more efficient than combining the individual crosscorrelations in this way.

2.2 Validation of the Revised Method

Given the changes that have been made to the model it is necessary to validate the revised model against a number of known measured data sets. This is done first with the original measured data of Lavandier & Culling (2010) to ensure the model remains as accurate for their data as it did when using their model. Then a number of measured data sets from the literature are modelled, with differing configurations of target and interferer, before finally modelling a set of data collected in three as yet unpublished experiments. The criterion used to ensure the accuracy of the model is the ability of it to predict the variation within each measured data set. Typically, the data sets measured SRTs for an anechoic target speech source with either one, two, or three anechoic interfering sources. Two sets of measured data (Bronkhorst & Plomp (1988) and Culling *et al* (2004)) looked at the effects of using either ILDs or ITDs only, whilst the data of Lavandier & Culling (2010) used an anechoic target but reverberant interferer, and the unpublished data uses both reverberant target and interferers. In each case, the input to the model had to be created in such a way as to replicate the sources used in the experiments to collect the measured data. This was either through generating BRIRs to the same source and room specifications as was used in the original experiments, or through using Head Related Transfer Functions in the purely anechoic configurations.

2.2.1 Lavandier & Culling (2010)

Lavandier & Culling (2010) measured speech reception thresholds for target speech in a speech-spectrum noise background using a 1-up/1-down adaptive threshold measure (Plomp & Mimpen 1979). Their target speech was sentences taken from the Harvard Sentence List (IEEE 1969), and was always anechoic, whilst they varied the level of interaural coherence of the noise interferer. This was done through modelling virtual rooms using |WAVE (Culling 1996), and varying the size of the room, the source distances within each room, the absorption coefficients of each of the rooms materials and the azimuthal separation of the sources. Their room modelling did not incorporate a head, rather the "ears" of the listener were modelled as two omnidirectional microphones, 18 cm apart, and 1.5 m from the floor.

In order to model the data, room impulse responses were generated using |WAVE in the same configurations as those used in Lavandier & Culling (2010), and were modelled using the revised method. The results are shown in figure 2.3. Note that in their paper Lavandier & Culling measured 16 conditions, but only 15 have been modelled, because their final condition had anechoic impulse responses for both target and interferer, but used different noise sources for the left and right channels of the interferer to give an interaural coherence lower than is possible through room modelling. As the revised prediction model works from the binaural room impulse responses it is blind to the use of independent noise sources and is unable to predict this final condition.

The revised model is able to predict the data of Lavandier & Culling (2010) to a high level of accuracy, with a correlation coefficient of 0.98 ($p < 0.0001$) and Root Mean Square Error (RMSE) of 0.37 dB. This marginally out performs the original model of Lavandier & Culling, where they obtained a correlation coefficient of 0.97.

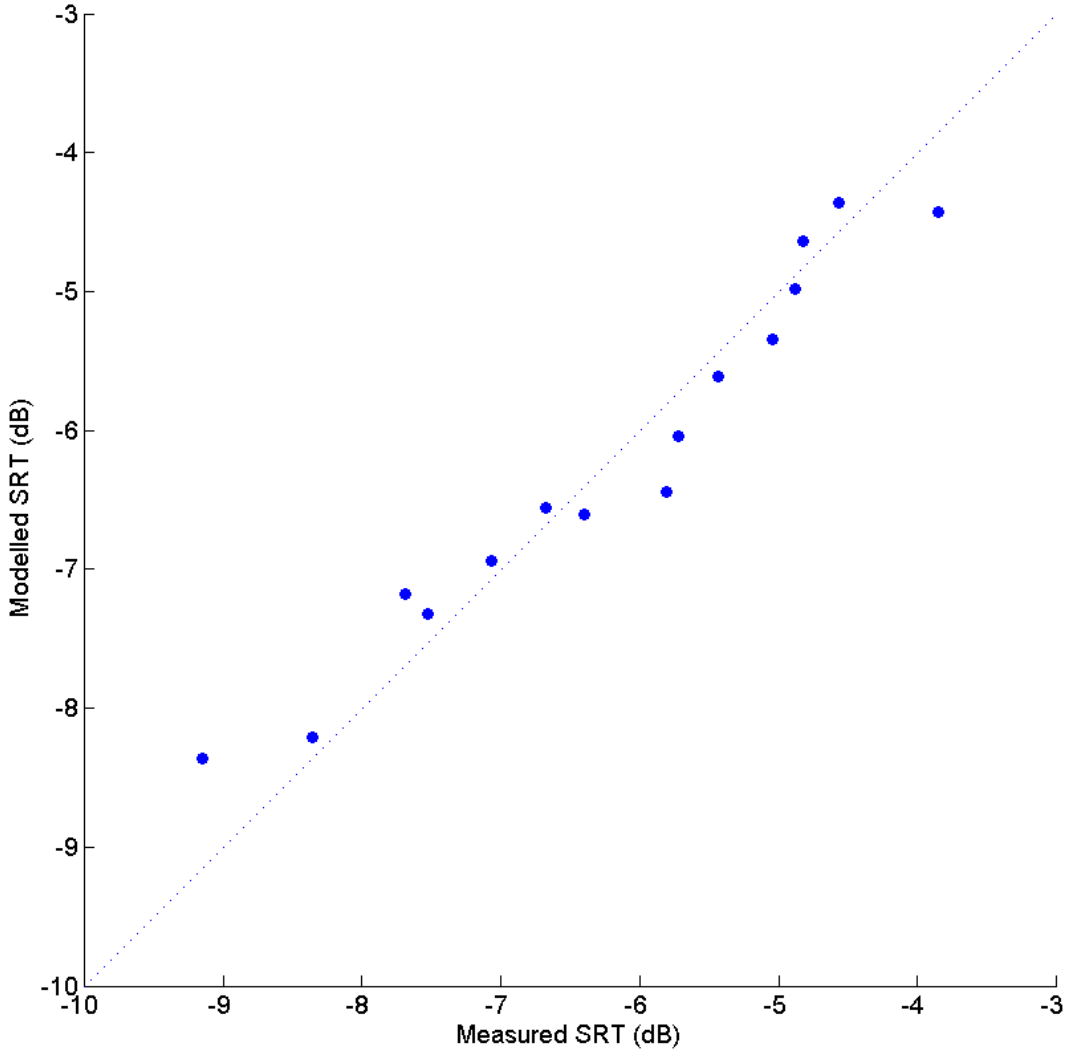


Figure 2.3: Results from modelling the data as measured by Lavandier & Culling (2010). Filled circles represent measured and predicted SRTs, dashed line is the 1:1 ratio line. r = 0.98, $p < 0.0001$. RMSE = 0.37 dB

2.2.2 Peissig & Kollmeier (1997)

Peissig & Kollmeier (1997) obtained speech reception thresholds for speech with either one, two, or three spatially separated noise interferers. They used a "subjective" adjustment method (Wesselkamp 1994, Kollmeier & Wesselkamp 1997) where listeners were allowed to alter the level of the target relative to the interferer to obtain a level which subjectively corresponded to 50% intelligibility. The test material was two sentences taken from the Göttinger Satztest, a German sentance intelligibility test (Wesselkamp, Kliem & Kollmeier 1992, Kollmeier & Wesselkamp 1997).

For each of the one, two, or three interfering conditions the target was presented at 0° azimuth, whilst one interferer was positioned at a range of 12 azimuths. In the two interferer condition a second fixed interferer was included at 105°, whilst in the three interferer conditions two fixed interferers were included at 105° and 255°. Peissing & Kollmeier simulated the different azimuthal angles by using real-time convolution of the noise signals with the outer ear impulse responses for the respective angles, taken from Pösselt *et al* (1986).

Modelling of the data was carried out using the revised method, and Head Related Transfer Functions (HRTFs) recordings of a KEMAR dummy-head from MIT (Gardner & Martin 1994). This is a significant change from the model of Lavandier & Culling (2010), as it now includes better-ear listening effects due to both head shadow and room colouration. As all of the conditions used in Peissig & Kollmeier (1997) are anechoic the HRTFs can be directly input into the model, with appropriate zero-padding, as conversion to binaural impulse responses would only entail convolution with a unit impulse, which would have no impact on the data extracted by the model. The results of the modelling can be seen in Figure 2.4, broken down by the number of interferers. The correlation between

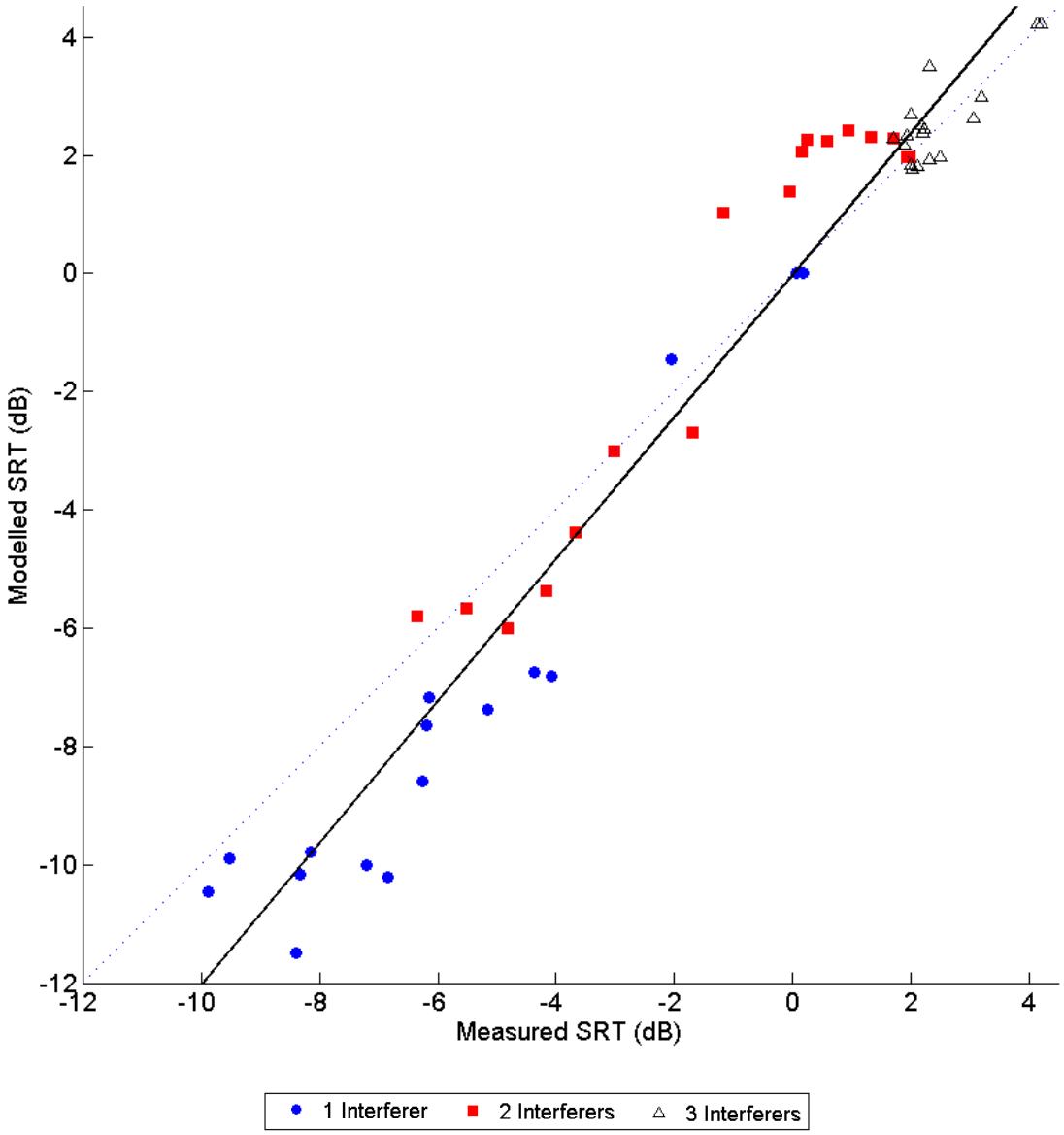


Figure 2.4: Results from modelling the data as measured by Peissig & Kollmeier (1997). Filled blue circles represent single interfering conditions, filled red squares represent two interferer conditions, and open black triangles represent three interferer conditions, the dashed line is the 1:1 ratio line. The solid black line is the linear regression across all three datasets, described by equation 2.3. r = 0.98, $p <$ 0.0001, RMSE = 1.58 dB.

the measured and the predicted SRTs is 0.98 ($p < 0.0001$) when comparing between all conditions, with an RMSE of 1.58 dB. Linear regression across all three datasets yields the solid black line shown in Figure 2.4, which can be described by the equation:

$$y = 1.2x - 0.04 \tag{2.3}$$

This systematic deviation from the desired 1:1 ratio is likely to be due to the paradigm used by Peissig & Kollmeier in the collection of their data. Bronkhorst (2000) proposed a formula for predicting the binaural release from masking for anechoic speech-in-noise configurations. As part of this formula there was a constant, C, which varied with data collection method and choice of speech material and acted as a scaling factor. For the data of Plomp & Mimpen (1981) and Bronkhorst & Plomp (1992) this constant could be set to 1, but for the data of Peissig & Kollmeier (1997) a value of 0.85 was required to get an accurate prediction. The proposed formula of Bronkhorst (2000) predicts the binaural gain due to spatial release from masking, whilst the revised model of Lavandier & Culling (2010) predicts the speech reception threshold. In essence these are the inversion of each other, and as such the scaling factor of 0.8 used by Bronkhorst, and the deviation from the 1:1 line by a factor of 1.2 in the predicted data of figure 2.4 are approximately equal.

2.2.3 Hawley *et al* (2004)

Hawley *et al* (2004) measured SRTs for Harvard IEEE sentences in a background of between one and three interferers. All sources were anechoic, with the target presented to the front of the listener at 0° azimuth, and interferers presented from azimuths of -30°, 0°, 30°, 60°, or 90°. The virtual stimuli were created using head-related impulse responses from the AUDIS database released by Blauert *et al* (1998). SRTs were measured for four

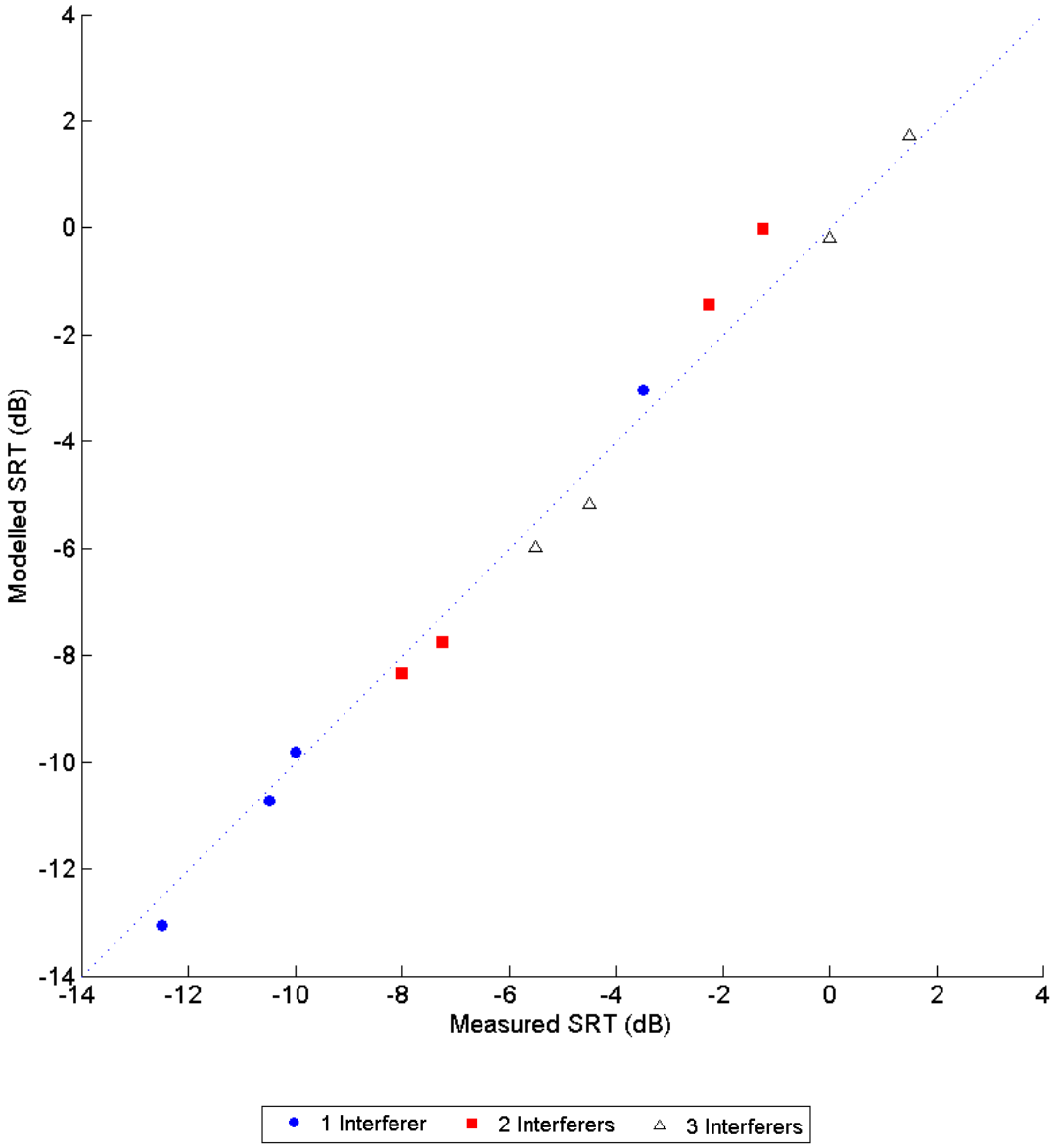


Figure 2.5: Results from modelling the data as measured by Hawley *et al* (2004). Filled blue circles represent single interfering conditions, filled red squares represent two interferer conditions, and open black triangles represent three interferer conditions, the dashed line is the 1:1 ratio line. r = 0.99, $p < 0.0001$, RMSE = 0.57 dB.

different interfering source types: (1) other sentences spoken by the same speaker as the target, (2) time-reversed sentences spoken by the same speaker as the target, (3) speech-spectrum shaped noise, and (4) speech spectrum shaped noise, modulated by the temporal envelope of the interfering sentences from (1).

Modelling was carried out on the data for the speech-shaped continuous noise interferers (3) only, using HRTFs from the MIT database of a KEMAR artificial head (Gardner & Martin 1994). The results are shown in Figure 2.5. The modelling results were normalised to the reference conditions used in Hawley *et al* (2004). The model was able to predict the measured data to a high level of accuracy, with a correlation coefficient of 0.99 ($p < 0.0001$) when comparing across one, two, and three interfering conditions, and an RMSE of 0.57 dB.

2.2.4 Bronkhorst & Plomp (1988)

Bronkhorst & Plomp (1988) measured speech reception thresholds for speech with a single noise interferer at a range of azimuths (0°, 30°, 60°, 90°, 120°, 150°, and 180°). HRTFs were generated by Bronkhorst & Plomp using a KEMAR manikin in an anechoic chamber. Noise samples were modified to allow for three different types of interferer, (1) the original noise sources, containing both ITD and ILD cues, termed FF (free-field), (2) samples with ILDs only, generated by modifying the temporal characteristics of the noise so as to match that of the 0° noise, termed dL, and (3) samples with ITDs only, generated by modifying the spectral characteristics of the noise so as to match that of the 0° noise, termed dT. These three interferer types allowed for investigation into either the better-ear effects only, using the dL interferers, the binaural unmasking effects only, using the dT interferers, or the combined effects, using the FF interferers.

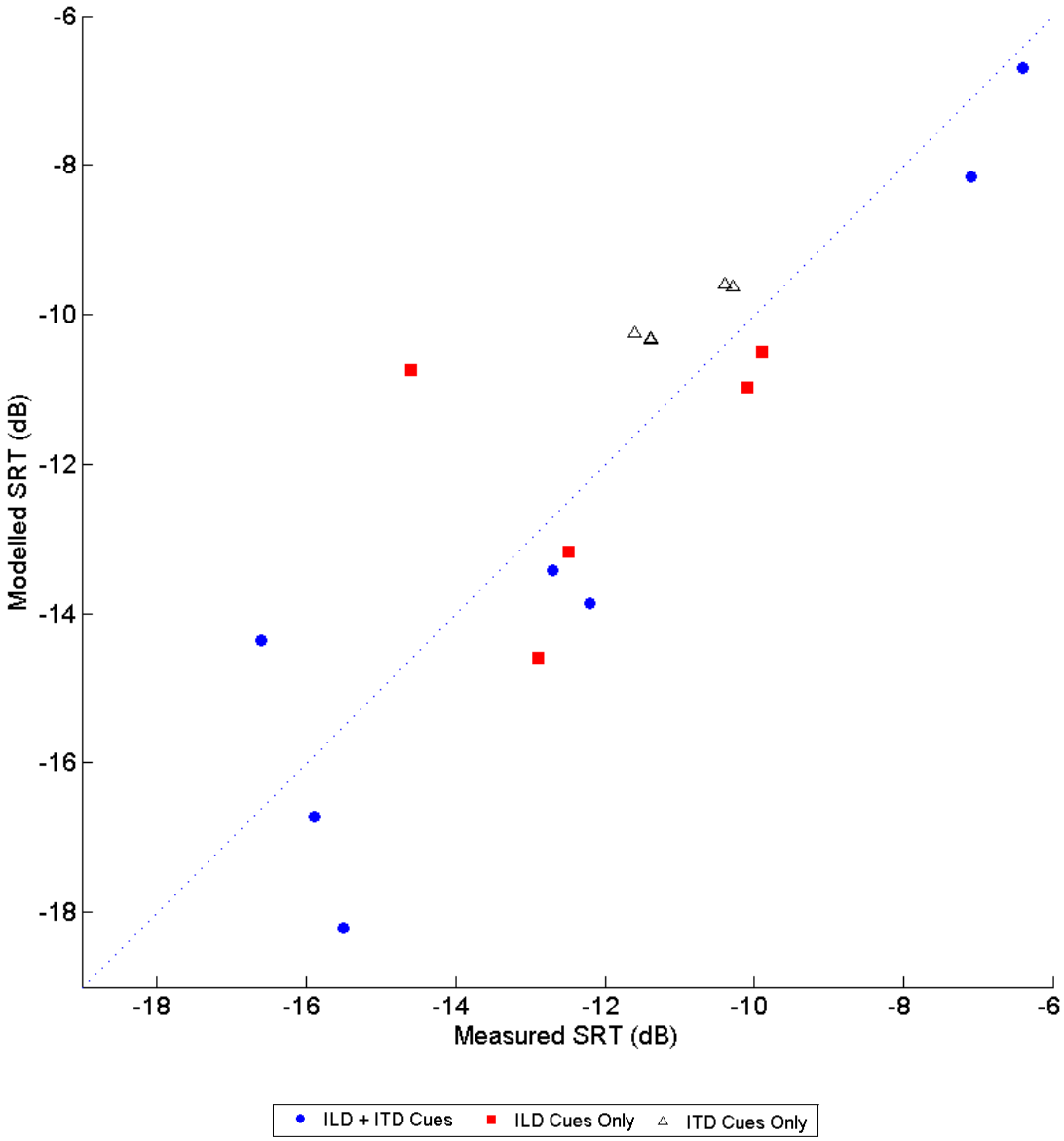


Figure 2.6: Results from modelling the data as measured by Bronkhorst & Plomp (1988). Filled blue circles represent the FF (free-field) conditions, with both ITD and ILD cues, filled red squares represent the dL conditions, with only ILD cues, and open black triangles represent the dT conditions with only ITD cues, the dashed line is the 1:1 ratio line. r = 0.86, $p < 0.0001$, RMSE = 1.57 dB.

As with the Peissig & Kollmeier (1997) and Hawley *et al* (2004) data, modelling was carried out using the HRTFs from the MIT database (Gardner & Martin 1994). To predict the dL data, only the values produced by the better-ear listening pathway of the model were considered. For the dT data only the values produced by the binaural unmasking pathway of the model were considered. Bronkhorst & Plomp (1988) did not collect data for the 0° and 180° interferer azimuths for the ILD- and ITD-only cases, as these should be the same as for the ILD and ITD combined case. Figure 2.6 shows the modelled data as a scatter plot for the ILD-only, ITD-only, and ILD & ITD combined conditions. The correlation coefficient for all of the three cue types combined is 0.86 ($p < 0.0001$), with an RMSE of 1.57 dB.

2.2.5 Culling *et al* (2004)

Culling *et al* (2004, 2005) measured SRTs for speech with three spatially distributed noise interferers. All sources were generated by convolving the material with the relevant anechoic HRTF recordings from the HMS III acoustic manikin from the AUDIS catalogue (Blauert, Brueggen, Bronkhorst, Drullman, Reynaud, Pellieux, Krebber & Sottek 1998). The interfering noises were processed in such a way that the end product was similar to that used by Bronkhorst & Plomp (1988), namely replacing either the phase or the amplitude data of the Fourier transform to give sources with either ILD cues only, or ITD cues only. They used the same azimuth separations for all cue types as those used by Hawley *et al* (2004) for their three-interferer conditions.

The results of the SRT measurements were modelled using the same impulse responses as were used by Culling *et al*, having undergone the same processing methods. The modelling results are shown in Figure 2.7, separated into ILD and ITD, ILD-only,

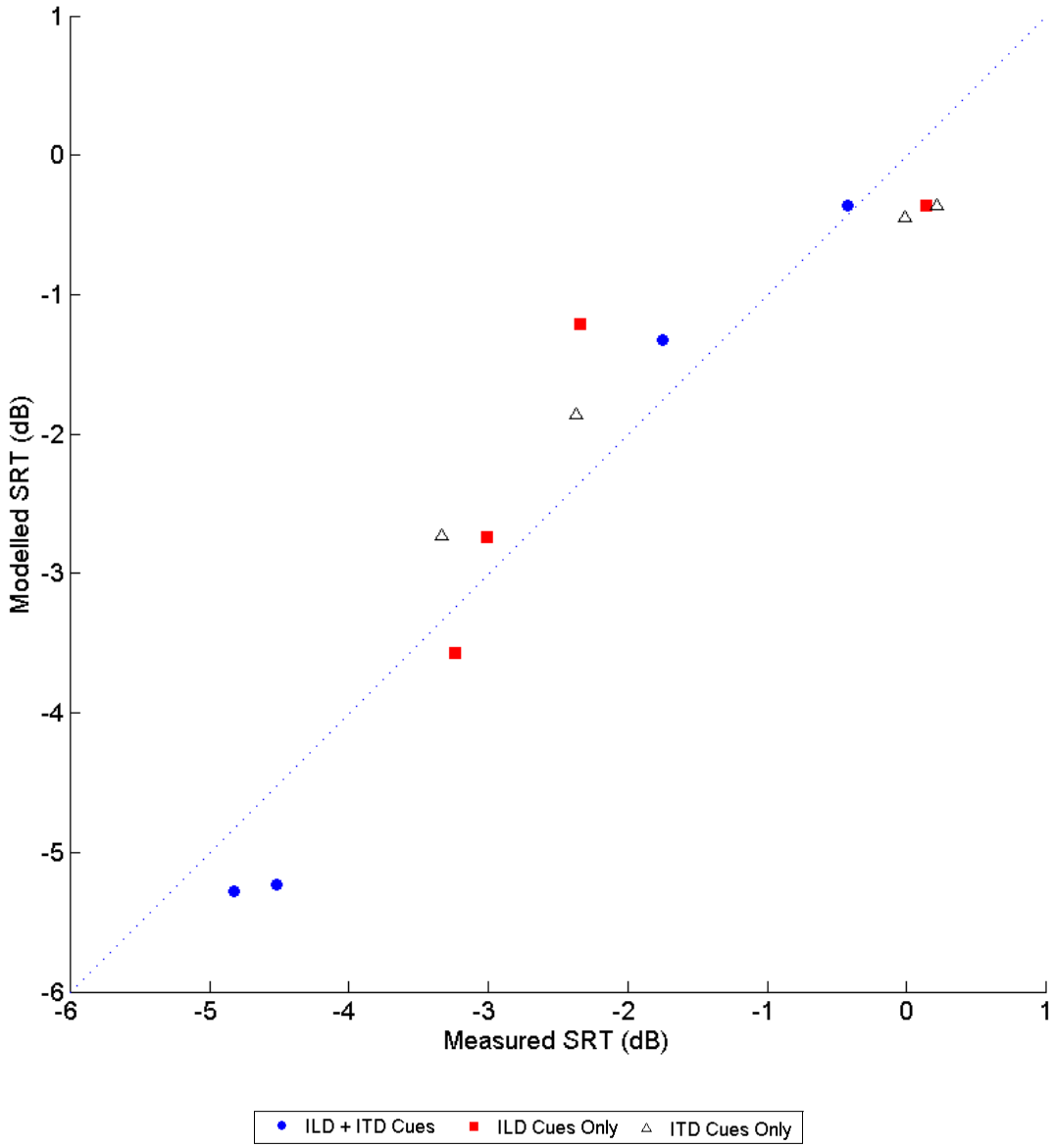


Figure 2.7: Results from modelling the data as measured by Culling *et al* (2004). Filled blue circles represent conditions with both ITD and ILD cues, filled red squares represent conditions with only ILD cues, and open black triangles represent conditions with only ITD cues, the dashed line is the 1:1 ratio line. r = 0.95, $p <$ 0.0001, RMSE = 0.56 dB.

and ITD-only conditions. The model is able to predict the measured data to a high level of accuracy, with a correlation coefficient of 0.95 ($p < 0.0001$) and an RMSE of 0.56 dB.

2.2.6 Unpublished Data

Lavandier & Culling (2010) measured SRTs for an anechoic target with a mixture of anechoic and simulated reverberant interferers. Three further experiments have been conducted by Dr Lavandier and Prof. Culling at Cardiff University, into the ability of listeners to understand speech in noise using real reverberation on both the target and the interferer.

BRIRs were supplied by Dr Anthony J. Watkins and colleagues at Reading University. These had been measured in five rooms using a Maximum Length Sequence technique in a corridor (240 m^3) and an L-shaped room (180 m^3) (Watkins 2005), and log sine sweeps (Farina 2000) in two meeting rooms (130 m^3 each) and a lecture hall (500 m^3). In all instances a KEMAR manikin was used for the listener, with a Bruel & Kjaer 4128 Head-and-Torso simulator as a source, with realistic directional characteristics.

In Experiment 1, SRTs were measured for a single interferer in one of the meeting rooms, with the target at 0.65 m from the listener at 25° azimuth. A single interferer either 0.65 m or 5 m from the listener was at either -25°, 0°, or 25° azimuth. The six interferer BRIRs used for these configurations were either processed to remove ITDs (and therefore the benefit of binaural unmasking), or were not, giving a total of 12 conditions.

In Experiment 2, 12 conditions were tested, again using a single interferer but using BRIRs from the four other rooms available. In the corridor the target was at 0.65 m and 0° azimuth, with the interferer either at 1.25 m or 2.5 m and 0° azimuth. In the L-shaped room, the target was in the same position, but the interferer distance increased to 5 m or

10 m. In the other meeting room, the target was at 25° to the left of the listener, with the interferer at 0.65 m, 1,25 m or 5 m from the listener and 25° azimuth. Finally, in the lecture theatre the target was either at 25° to the left of the listener, whilst the interferer was at 25° and either 0.65 m, 2.5 m, or 10 m, or the target was to the right of the listener whilst the interferer was to the left at 0.65 m. In this experiment no processing was done of the BRIRs.

In the final experiment of this series, Experiment 3, the meeting room from experiment 1 was used again, but this time with either 1, 2, or 3 interferers, at 0.65 m or 5 m from the listener, with a range of azimuths.

Appendix A (p. 7.4) contains tables showing the configurations of all of the sources used in these three experiments.

The results of modelling the three laboratory experiments are shown in figure 2.8. The model is able to predict the measured results of all three experiments to a high degree of accuracy, with correlation coefficients of 0.98, 0.95, and 0.98, and RMSEs of 0.32, 0.35, and 0.49 dB for experiments 1, 2, and 3 respectively ($p < 0.0001$ in all cases).

2.3 Discussion

Lavandier & Culling (2010) showed that their original model was able to predict SRTs for anechoic target speech with a variety of reverberant single interferers. Their data can be used to validate the revised model, ensuring that it is able to take into account the importance of both azimuthal separation of the sources, as well as the importance of interferer interaural coherence. After revision the model is still able to predict their findings to a high degree of accuracy, whilst making significant savings with respect to computational overheads, speed increases were possible in the order of 170x faster (2.7

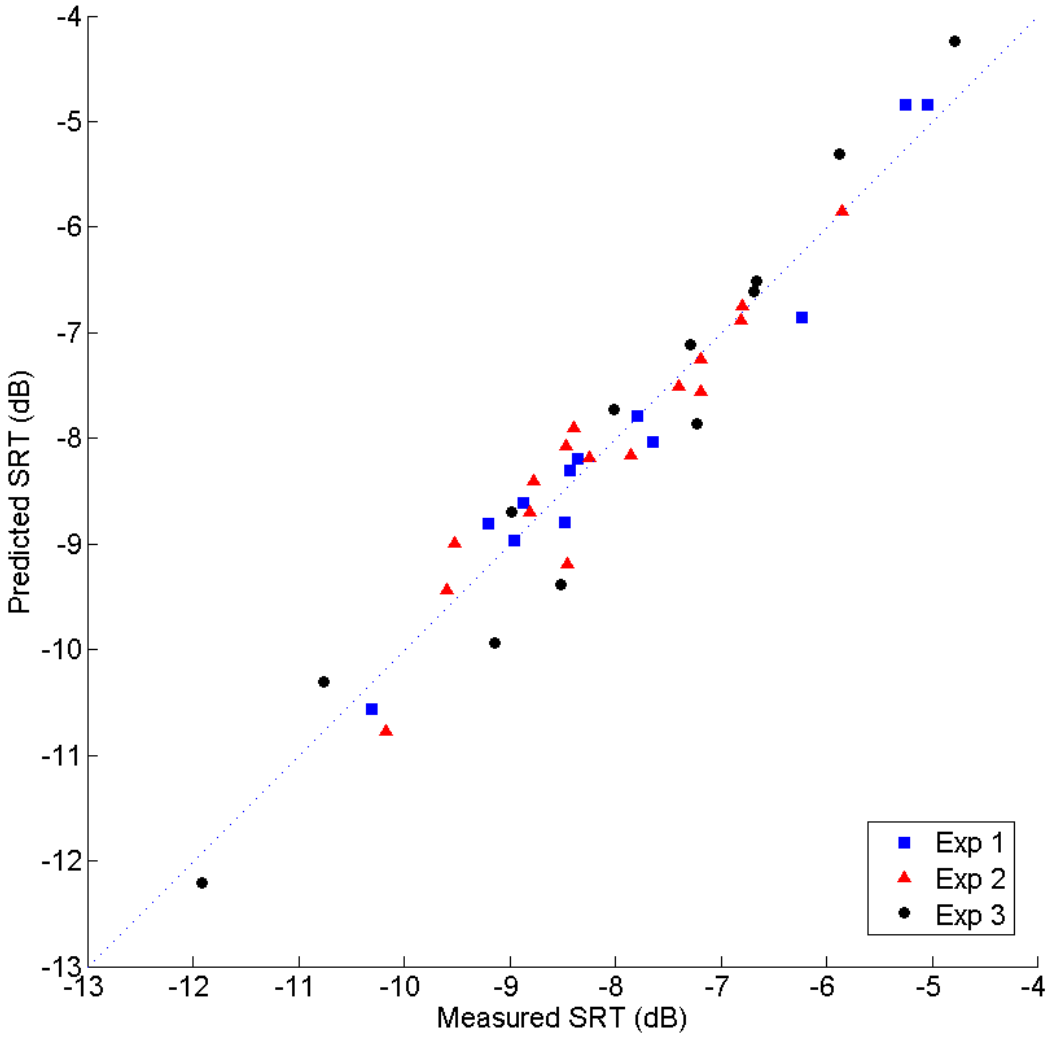


Figure 2.8: Results of modelling the data collected within the laboratory. Filled blue squares represent the data of Experiment 1 (r = 0.98, $p < 0.0001$), red triangles the data of Experiment 2 (r = 0.95, $p < 0.0001$), and black circles the data of Experiment 3 (r = 0.98, $p < 0.0001$). The dashed line is the 1:1 ratio line. The RMSE for the three experiemtns is 0.32, 0.35, and 0.49 dB respectively.

seconds per prediction compared to 7 minutes, 40 seconds per prediction for the model of Lavandier & Culling (2010)).

Through modelling of the data of Peissig & Kollmeier (1997) and Hawley *et al* (2004) it can be demonstrated that the model is able to accurately handle listening situations with multiple spatially separated interferers, even when differing head related impulse responses / head related transfer functions are used in the collection of the data to those used in the modelling of it. In the case of Peissig & Kollmeier (1997), the model also accurately predicted SRTs collected using a non-English language.

The model was less able to predict the results of Bronkhorst & Plomp (1988). Looking at Figure 2.9 it can be seen that the principal point that is predicted poorly is in the ILD-only set, with the interferer at 90° to the listener. At this point the model predicts a drop in intelligibility (a higher SRT) which is not evident in the data. The reason that the model predicts this is that with a source at 90° to the listener the ILD is smaller than at other angles. This occurs because sound travelling around both sides of the head have roughly equal path lengths to the contralateral ear, and so arrive in phase. This causes constructive interference, which in turn raises the sound pressure level, and reduces the effectiveness of head shadowing, creating a "bright spot". This effect was first observed in acoustics by Lord Rayleigh (1904), but also much earlier in optics where it is referred to as the Arago Spot, or Poisson Spot.

Figure 2.10 replicates Figure 2 of Bronkhorst & Plomp (1988), showing the 1/6th Octave normalised response curves of a KEMAR manikin at a range of azimuths directly replicated from the paper as Bronkhorst & Plomp presented them (Thick lines, Solid - ipsilateral ear, Dashed - contralateral ear), with the addition of the equivalent curves as derived from the Gardner & Martin (1994) recordings of the HRTFs of a KEMAR manikin

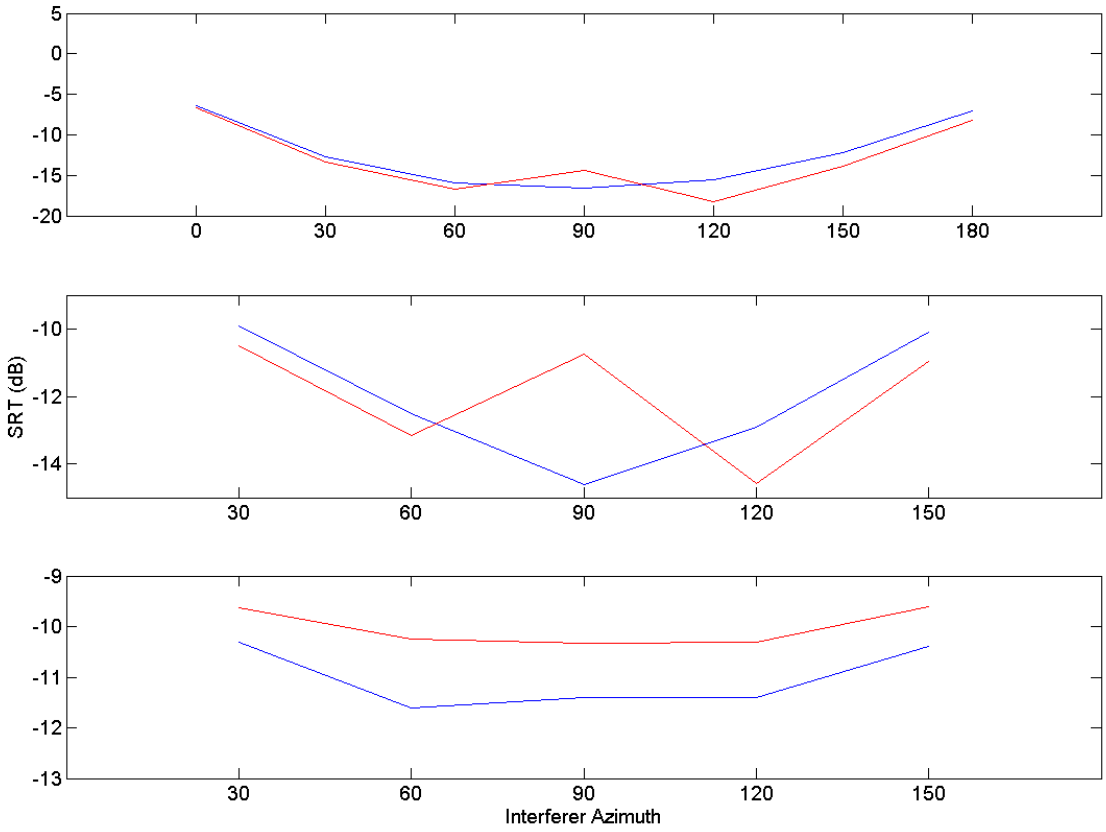


Figure 2.9: Results from modelling the data as measured by Bronkhorst & Plomp (1988) as a factor of interferer azimuth, blue lines represent the measured data, red lines the modelled data. The top plot shows the conditions with both ILD and ITD cues (r = 0.93, $p < 0.005$), the middle plot the conditions with ILD cues only (r = 0.33, $p > 0.5$), and the bottom plot the conditions with ITD cues only (r = 0.85, $p < 0.005$).

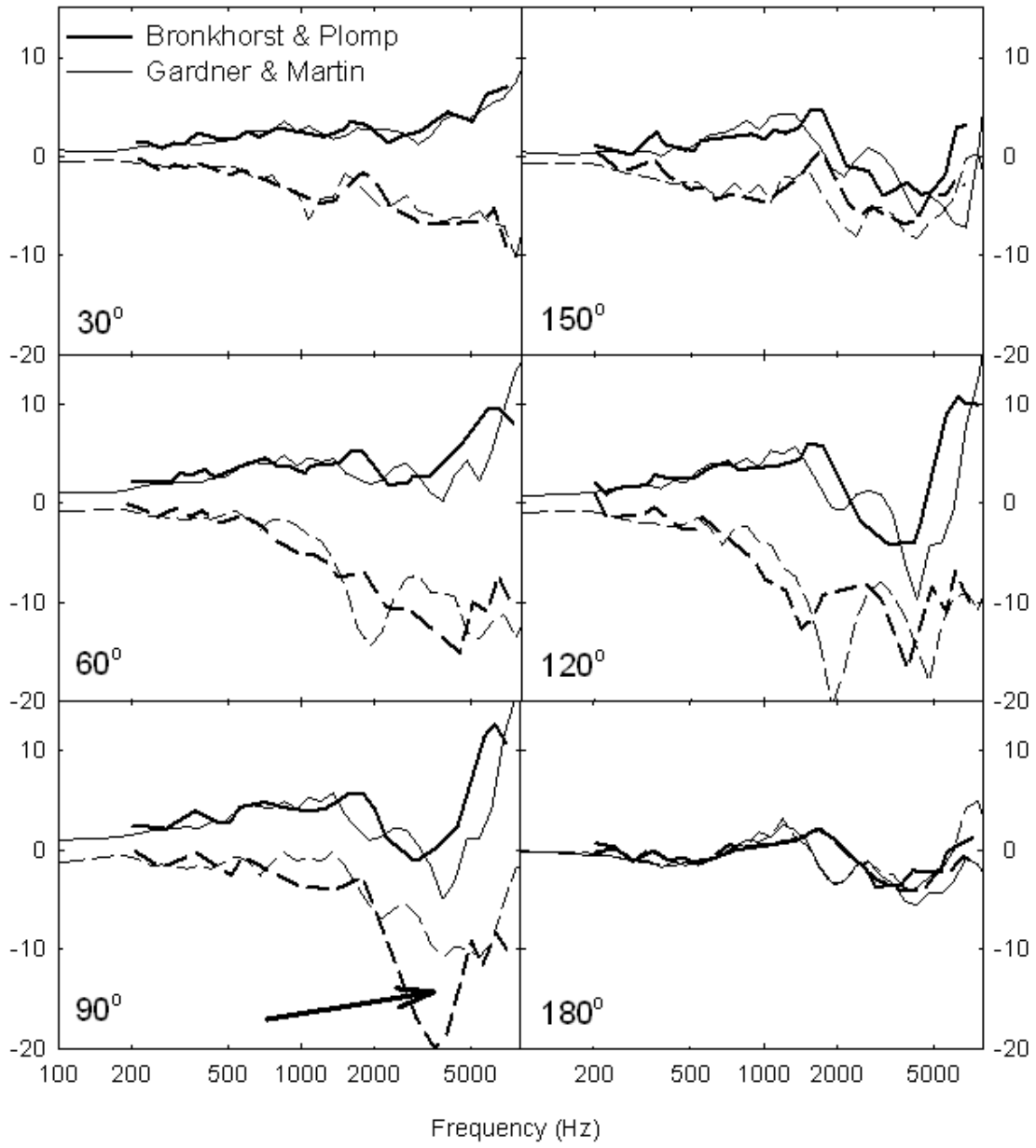
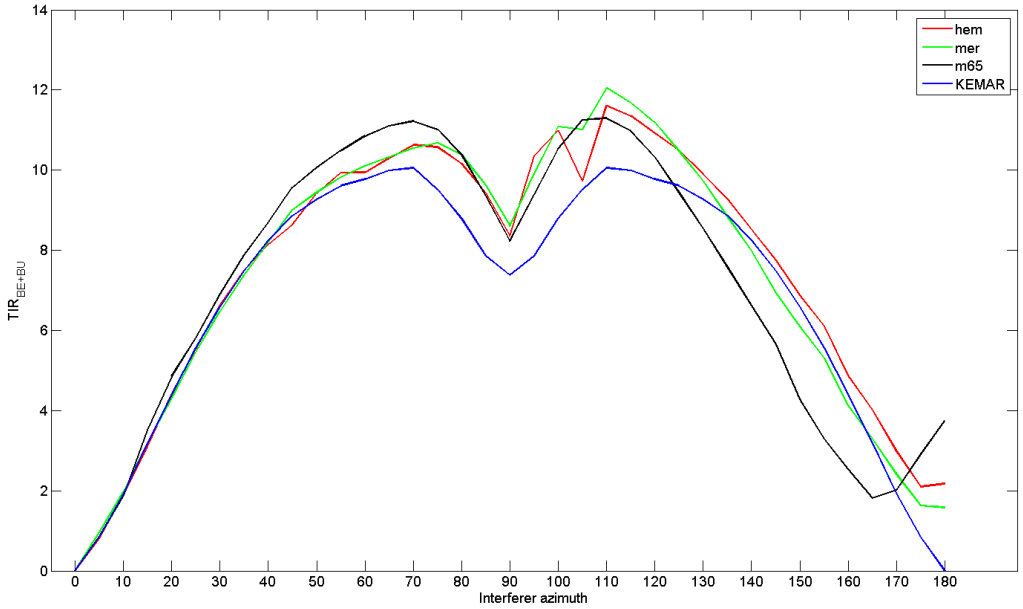


Figure 2.10: Replication of Figure 2 of Bronkhorst & Plomp (1988), Showing the normalised response curves of their KEMAR manikin, and the comparable curves of a KEMAR manikin based on measurements by Gardner & Martin (1994). Horizontal axis shows frequency, vertical axis is level, for a range of source azimuths.

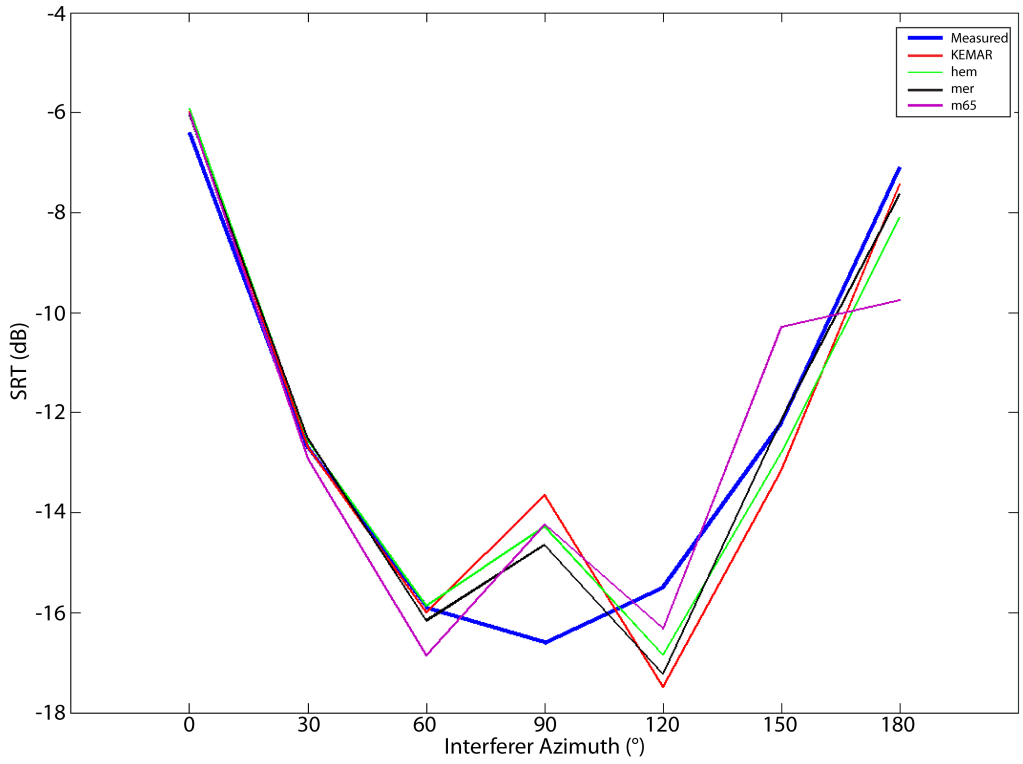
(Thin lines, Solid - ipsilateral ear, Dashed - contralateral ear). The arrowed area in the 90° condition highlights the difference in the contralateral ear between the HRTFs used by Bronkhorst & Plomp (1988), and those of Gardner & Martin (1994), as used in the present modelling. Note that these are recordings made from the same manikin type, but carried out in the two different laboratories. In this area Gardner & Martin's data predicts a rise in interferer intensity at the contralateral ear, when compared to the 60° condition, but Bronkhorst & Plomp (1988) do not observe this rise, but rather a reduction in intensity compared to the 60° equivalent. This difference in response between the two sets of HRTF measurements from the same manikin type will go some way to explain the differences in the measured and the modelled data, particularly in the 90° conditions. If the HRTFs used by Bronkhorst & Plomp (1988) were available it would likely be possible to predict their data to a much higher level of accuracy.

One concern is that due to the high symmetry of a KEMAR manikin, coupled with the fact that the Gardner & Martin (1994) HRTFs are perfectly symmetrical (i.e. the left ear $+10^\circ$ HRTF and the right ear -10° HRTF are identical) the prominence of the "bright-spot" effect may be over predicted. To ascertain whether this was the case modelling was carried out using the Gardner & Martin KEMAR HRTFs, as well as with three sets of HRTF measurements of human heads taken from the AUDIS database (Blauert et al. 1998). The human HRTFs used were those of person "m65", measured at the TNO Human Factors Research Institute, and of persons "hem" and "mer", measured at the Institut für Kommunikationsakustik, Ruhr-Universität Bochum.

Figure 2.11(a) shows the effect of interferer azimuth on the predicted $\mathrm{TIR}_{\mathrm{BE+BU}}$, with a target source located at 0°, for the four different HRTF sets. In each case the "bright-spot" effect is clearly visible, with the HRTFs of person m65 having a very similar pattern as to



(a) Effect of interferer azimuth on predicted TIR_{BE}, for 4 different sets of HRTFs. Target always located at $0°$. All sets modelled with $5°$ increments, but symbols represent measured HRTFs, data between symbols uses interpolated HRTFs.



(b) Replication of the modelling for the Bronkhorst & Plomp (1988) "ff" condition, top panel of fig. 2.9, with differing HRTFs. r = 0.93, 0.96, 0.97 and 0.91, and RMSE = 1.58, 1.26, 1.25 and 1.82 dB for KEMAR, hem, mer, and m65 respectively,

Figure 2.11: Effect of interferer azimuth on predicted TIR_{BE+BU} for 4 differing HRTF sets, and replication of modelling the Bronkhorst & Plomp (1988) "ff" condition with differing HRTFs.

that of a KEMAR manikin. Whilst the overall predicted level for person m65 is greater, the range of azimuths over which the bright-spot effect occurs, and the relative decrease in predicted $\text{TIR}_{\text{BE+BU}}$ due to the bright spot are both in-keeping with the pattern as predicted by the KEMAR manikin. Due to the inherent difficulties of measuring human HRTFs, such as getting the alignment of the head relative to the source correct, there is some experimental variation included within the plots. Also, the human HRTFs are measured at 15° increments (as denoted by the symbols on figure 2.11(a)), with the HRTFs at other azimuths being interpolated from those that are measured. The CIPIC database of human HRTFs (Algazi, Duda, Thompson & Avendano 2001) could have been used instead of the AUDIS catalogue, but whilst the CIPIC database measures the HRTFs at 5° increments between -45° and $+45^\circ$ azimuth, which would give a clearer picture of what is happening to the front of the listener, outside of this range they are measured at only $\pm 55^\circ$, $\pm 65^\circ$, and $\pm 80^\circ$, crucially missing the $\pm 90^\circ$ position that is of interest for this modelling.

Figure 2.11(b) replicates the "ff" condition of Bronkhorst & Plomp (1988), as shown in the top panel of fig. 2.9. In this case the four differing sets of HRTFs have been modelled, with all of them giving broadly the same prediction in SRT, and all of them including the rise in SRT with the interferer at 90° that is not prevalent in the measured data. This reinforces the assumption that the significant difference in prediction for this one configuration is due in some way to the measurement and processing of the original HRTFs used by Bronkhorst & Plomp in the generation of their test material.

Though the accuracy of the model was reduced when predicting the data of Bronkhorst & Plomp (1988), it predicted the data of Culling *et al* (2004) to a high level of accuracy. Culling *et al*'s experiment was similar to that of Bronkhorst & Plomp's, in that it included both the ITD- and ILD-only conditions. Due to artefacts introduced by Culling *et al*

(2004) through their processing techniques, it was necessary to use the same set of impulse responses processed using the same techniques as the input to the model. Different impulses were thus required for the ILD- and ITD-only conditions. This was different to the modelling method used for Bronkhorst & Plomp's data, for which the result of either the binaural unmasking, better-ear listening or their combination was taken from the model output for the same impulse responses. The effect of the processing used by Culling *et al* (2004) can be seen further from the fact that the results obtained by them for the ILD and ITD combined cases is different from the results obtained by Hawley *et al* (2004) in their three interferer conditions. It would be expected that the data of Culling *et al* (2004) and Hawley *et al* (2004) would be the same for these conditions given that they are using the same methods and materials, configured in the same way. This is important as it shows that small differences in the generation of the input sources can have significant effects on the modelled results. Modelling of the Culling *et al* data with the revised model demonstrates that it is able to predict not only the combined effects, but also the individual contributions of binaural unmasking and better-ear listening. The model proposed by Lavandier & Culling (2010) was only demonstrated to be able to predict better-ear listening due to room colouration, whilst the revised model can be shown to account for both better-ear listening due to room-colouration and due to head shadow.

By modelling a number of different data sets it has been possible to validate the model, showing that it can accurately predict the importance of both binaural unmasking and better-ear listening in situations that have one, two, or three interferers in either an anechoic, simulated reverberant, and real reverberant listening environment.

Chapter 3

Example Room Maps

One of the major benefits of the revised model is that its increased speed allows for the calculation and prediction of many configurations in a short period of time. It can thus more effectively map many receiver positions within a room for any configuration of target and interferer sources. This chapter explores some of the potential of this technique for generating maps of intelligibility levels in rooms from their design.

As shown previously, listeners use both better-ear listening and binaural-unmasking to aid the separation of target speech from a background of interfering noise sources. For situations where listening takes place in rooms there are two key questions to be answered by this chapter; what is the effect of reverberation on the listeners' ability to understand speech, and what is the effect of multiple interferers being present on the listeners ability to understand speech? Whilst the existing models show the effects of reverberation on the target speech signal (i.e. the Speech Intelligibility Index (ANSI 1997), and the Speech Transmission Index (BSI 2003)), they do not include the effects of reverberation on the perceptual separation of the interferer, nor the effects of the number or spatial distribution of interfering sound sources.

The method of generating the intelligibility maps has two main processes; the first is the generation of the source-receiver Impulse Responses (IRs) using architectural acoustic software, and the second is the processing of these impulse responses using the revised psychoacoustic model described in Chapter 2. By altering the architectural acoustic model it is possible to either individually include or exclude the effects of head-shadow, azimuthal separation of sources, and reverberation, whilst alteration of the psychoacoustic model allows for either inclusion or exclusion of the effects of better-ear listening and of binaural-unmasking. By altering both the architectural and psychoacoustic model it is possible to generate a number of maps, principally, but not limited to:

1. Omnidirectional Target-to-Interferer Ratio (TIR_{OMNI}).

   Omnidirectional maps represent a reference case. They consider speech reception through a single omnidirectional microphone with no head present. This shows the simple effect of reverberation and distance on the target-to-interferer ratio at the listening location, and excludes all aspects of binaural and spatial hearing.

2. Effective Target-to-Interferer ratio from better-ear listening (TIR_{BE}).

   The effective target-to-interferer ratio with better-ear listening is modelled by including the effect of the head in the frequency response at each ear. This is a combination of room colouration, included in the IRs in the reverberant situations, and head shadow, in both anechoic and reverberant situations. The effects of head shadow are predicted by including the Head Related Transfer Functions (HRTFs) from a KEMAR manikin, as measured by Gardner & Martin (1994). The Target-to-Interferer ratio from better-ear listening is calculated using only the better-ear pathway of the revised model, as explained in chapter two.

3. The predicted benefit due to better-ear listening (B_{BE}).

   The benefit due to better-ear listening is the difference between the first two maps, the omnidirectional target-to-interferer ratio (1) and the target-to-interferer from better-ear listening ratio (2).

4. The target-to-interferer ratio with better-ear listening and binaural unmasking (TIR_{BE+BU}).

   The combined better-ear listening and binaural-unmasking map is calculated using the same IRs as those used for the better-ear listening only map, with the effects of room colouration and head shadow, but uses both the better-ear and binaural unmasking pathways of the revised model. (See figure 2.1(a), Page 23)

5. The predicted benefit due to binaural-unmasking (B_{BU}).

   The benefit due to binaural-unmasking is calculated as the difference between the combined map (4) and the better-ear listening only map (2).

As discussed in Section 2.1.1 (p. 21), the benefits as calculated by the model are additive, that is, the predicted TIR_{BE+BU} is equal to the predicted TIR_{OMNI} plus the B_{BE} and B_{BU}.

To explore the effects of reverberation and multiple interferers, two configurations were modelled, with either a single interferer or with three interfering sources, in both a reference anechoic space and in a typical reverberant space. In all cases, the space measured 6.4 m by 10 m. The target and interfering sources remained stationary, whilst the listener was modelled across a 0.3 m by 0.3 m square grid at a height of 1.5 m, which gives 693 points across the grid. For each position, where the head was included, the head was orientated so that the listener was facing the target and all BRIRs were generated

using an adapted version of the room impulse response program in |WAVE (Culling 1996), which filters each echo in the room impulse response by the HRTF corresponding to the direction of the incident ray. The HRTFs as measured by Gardner % Martin (1994) are at a minimum of 5° steps in the horizontal plane, so incident rays are quantised into corresponding incident angles. The target and interfering sources were assumed to be omnidirectional loudspeakers.

For each configuration, be it with single or multiple interferers, in an anechoic or a reverberant space, all five of the above maps were generated and analysed. Note, for all maps where a head is included, that is every map other than the TIR_{OMNI} maps, the listener is orientated so as to always face the target.

3.1 Single-Interferer Configurations

For the single interferer configurations a target was located at 4.4 m by 5.5 m, with an interferer located at 3.9 m by 2 m within the 6.4 m by 10 m listening environment. In the anechoic case, the absorption coefficient of all the surfaces in the room was set to be 1 at all frequencies, whilst in the reverberant case all of the walls were set at 0.4, with the ceiling at 0.9 and the floor at 0.2 at all frequencies. The target and interferer had the same relative power.

This set of absorption coefficients was chosen so as to be representative of a mildly reverberant typical office environment, with mildly absorptive walls, carpet on a concrete floor, and an acoustic suspended ceiling. The upshot of this configuration is that it accentuates the lateral reflections whilst reducing the reflections from the ceiling. First-order reflections from the ceiling and floor may arrive at the listener with the same azimuthal angle but differing elevation to the direct sound, this will tend to increase the level of inter-

aural coherence, which in turn may alter the effect of reverberation on binaural-unmasking and better-ear listening, without altering the statistical reverberation time.

This configuration of absorption coefficients gives a statistical reverberation time of 0.3 seconds, as calculated using the Sabine equation:

$$RT_{60} = \frac{0.161V}{\sum S_i \alpha_i(f)} \tag{3.1}$$

where V is the volume of the room in m^3, S is the area of each surface in m^2, and $\alpha(f)$ is the frequency-dependent absorption coefficient of each surface.

3.1.1 Anechoic Space

The first configuration modelled is that of a single interferer in an anechoic room. This is the simplest of the configurations, and allows investigation of the effect of the relative target and interferer azimuthal angles in relation to the listener, and the effect of the target and interferer distances from the listener.

3.1.1.1 Omnidirectional map

Figure 3.1 shows the results of modelling the TIR_{OMNI} for a single microphone in space. The solid line denotes the 0 dB contour line, with the dashed and dotted lines representing the ± 1 dB and ± 3 dB contours respectively.

As this configuration is an anechoic space with omnidirectional receivers the target-to-interferer ratio is governed solely by the distance of the target and interferer to the listener, and as such is inversely-proportional to the distance. This gives rise to the perfectly straight 0-dB contour line perpendicular to the target-interferer axis, as this is where the listener distance is equal to both sources, with the pattern being symmetrical about the

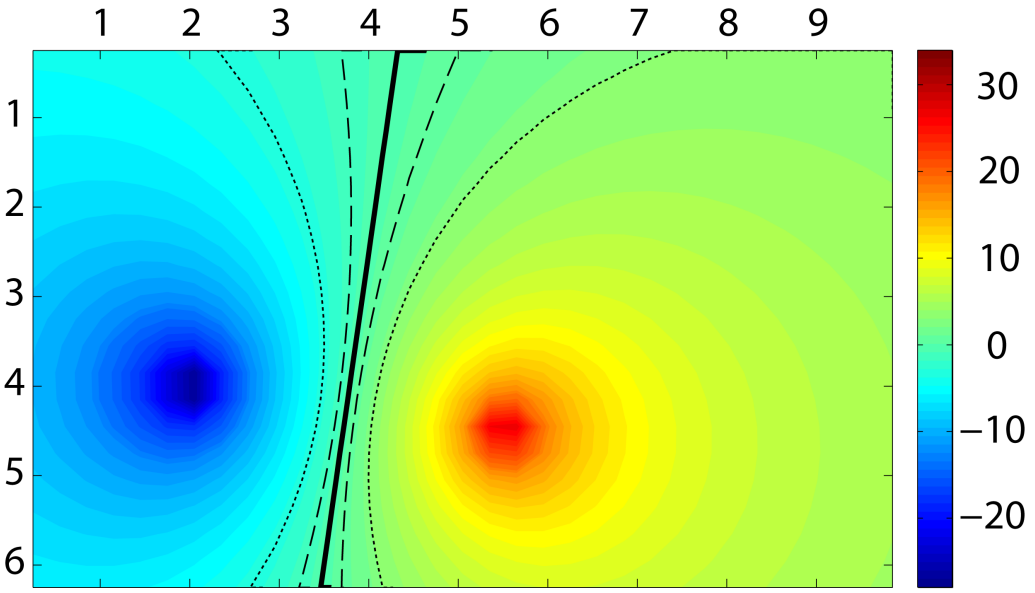


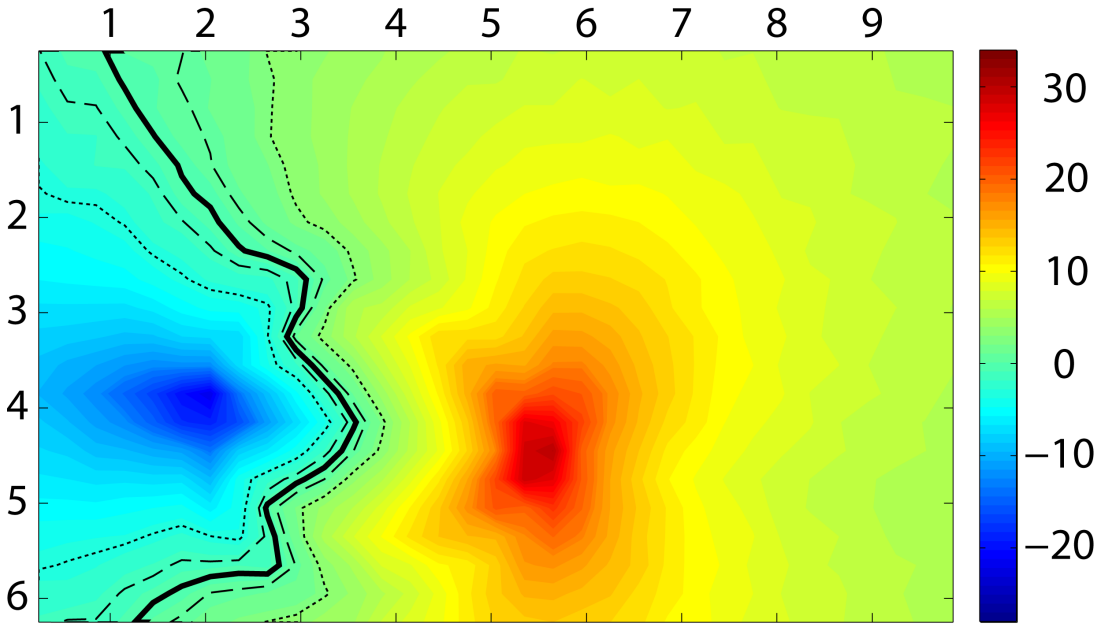
Figure 3.1: Map of TIR_{OMNI} in an anechoic 6.4 m x 10 m room, single interferer only, Target at 4.4 m by 5.5 m, Interferer at 3.9 m by 2 m.

target-interferer axis. Table A.4 (p. 189) shows the mean and standard deviation values for this, and all other Target-to-Interferer Ratio (TIR) maps.

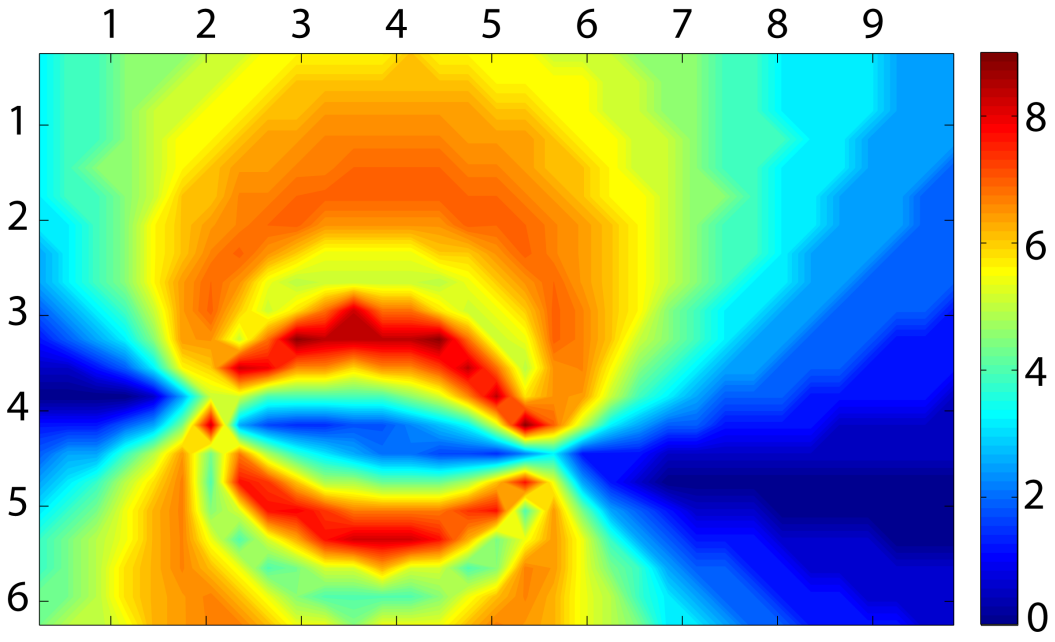
3.1.1.2 Better-Ear Listening Only.

Figure 3.2(a) shows the modelled effective TIR_{BE} and Figure 3.2(b) shows the B_{BE} (Fig. 3.2(a) minus Fig. 3.1). From Figure 3.2(a) we can see the overall increase in TIR, with the 0 dB (solid) contour line being pushed closer to the interferer, indicating that more of the listening positions within the room have a positive effective TIR. This is confirmed by the statistics, with an increase in mean TIR from 1.2 dB in the omnidirectional case to 5.4 dB in the better-ear listening case (from table A.4).

Figure 3.2(b) shows the pattern of this benefit. Better-ear listening gives an increase in benefit in all areas where the listener is not on the target-interferer axis. Off axis the signals received by the two ears are different, with one ear receiving a higher target-to-interferer ratio than the other ear due to shadowing by the head. This effect of head shadow increases with greater azimuthal separation of sound sources, being greatest when the azimuthal separation is larger than 90°, though decreasing rapidly as the separation approaches 180°. One thing to note is the approximately circular dip in benefit, passing through the target and interferer locations. This is caused by the fact that at these locations, with the listener facing the target, the azimuthal separation between the target and the interferer is 90°. In this situation, the effectiveness of head-shadow and better-ear listening is reduced due to the "bright spot" caused by constructive interference between waves passing around different sides of the listeners head (Rayleigh 1904) which increases the level of the interferer received at the contralateral ("shadowed") ear. (See Chapter 2.3, p. 42 for further explanation of the bright spot.)



(a) TIR_{BE}



(b) B_{BE}

Figure 3.2: Maps of (a) TIR_{BE} and (b) B_{BE} in an anechoic 6.4 m x 10 m room, single interferer only, Target at 4.4 m by 5.5 m, Interferer at 3.9 m by 2 m. The listener is always facing the target.

3.1.1.3 Binaural Unmasking and Better-Ear Listening.

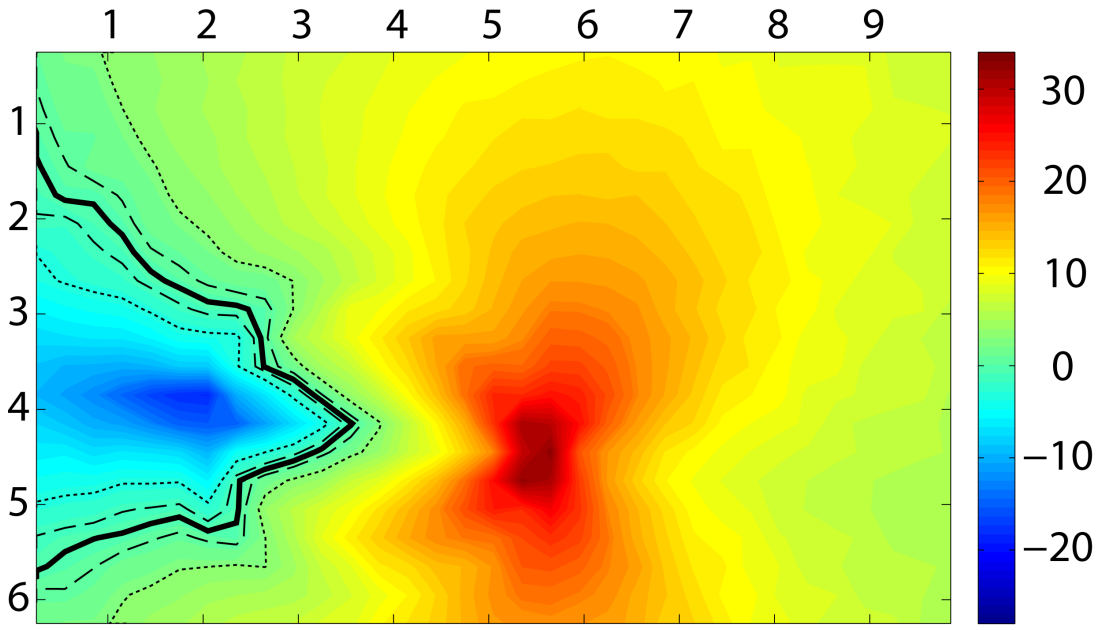
Figures 3.3(a) and 3.3(b) show the TIR_{BE+BU}, and the B_{BU} (Fig. 3.3(a) minus Fig. 3.2(a)) respectively. Comparing Figure 3.3(a) to the better-ear listening only map (Fig. 3.2(a)) there is an overall increase in effective TIR, with the 0 dB line being pushed even closer to the interferer; the mean has also been increased by 2.7 dB.

Figure 3.3(b) shows the pattern of this increase over better-ear listening alone, the benefit of binaural unmasking. As with the increase due to better-ear listening (Fig. 3.2(b)) the increase due to binaural unmasking occurs at all locations other than on the target-interferer axis. This is due to the requirement for interaural phase differences between the target and interferer. On the target-interferer axis both the target and interferer are at the same azimuth to the listener (0°), or directly in front and behind the listener (0° and 180°), and as such have the same interaural phase. As the listener moves away from this axis the level of binaural-unmasking increases, to a maximum of 3.8 dB.

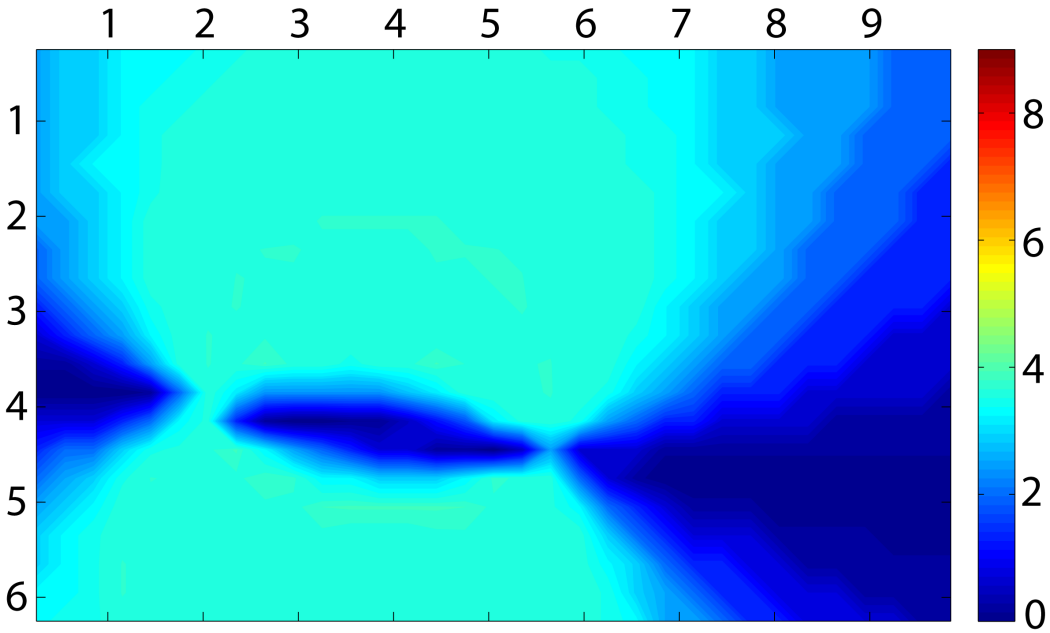
The maximal B_{BE} and B_{BU} are 8.8 dB and 3.8 dB respectively, but these occur at differing locations, resulting in an increase in the predicted maximum of the omnidirectional modelling map (Fig. 3.1) to the combined effects map (Fig. 3.3(a)) of 5.8 dB, and an increase in the predicted mean of 6.9 dB.

3.1.2 Reverberant Space

The second main configuration modelled is that of the single interferer in a reverberant space. This follows the same method and locations used in the anechoic space, but with a change in the absorption coefficients used for the surfaces. This allows for a comparison not only between the different types of map produced by the model, but also between the corresponding maps of the anechoic and the reverberant conditions, illustrating the effect



(a) TIR_{BE+BU}



(b) B_{BU}

Figure 3.3: Map of (a) TIR_{BE+BU} and (b) B_{BU} in an anechoic 6.4 m x 10 m room, single interferer only, Target at 4.4 m by 5.5 m, Interferer at 3.9 m by 2 m. The listener is always facing the target.

of reverberation.

3.1.2.1 Omnidirectional map.

Figure 3.4 shows the results of the omnidirectional modelling of a single microphone in the reverberant room. Compared to Figure 3.1 it can be seen that the 0 dB line is in approximately the same place, but that the ± 1 and ± 3 dB lines are spaced further apart, indicating that the rate of change across the space is decreased. The mean TIR changes from 1.2 dB in the anechoic condition to 0.5 dB in the reverberant condition. The pattern produced by the reverberant omnidirectional mapping is thus similar to that produced by the anechoic omnidirectional mapping, but with a reduction in the standard deviation of the values obtained. This is because reverberation equalises the sound levels by creating a diffuse field. The erratic effects of room colouration on the target and interferer spectra causes the iso-TIR contour lines to be less smooth and regulated.

3.1.2.2 Better-Ear Listening only.

Figures 3.5(a) and 3.5(b) show the effective $\mathrm{TIR_{BE}}$, and the $\mathrm{B_{BE}}$ respectively. Figure 3.5(a) shows that the effect of better-ear listening is to push the 0-dB contour line closer to the interferer, though not to the same extent as in the anechoic case (Fig. 3.2(a)). This is further illustrated by Figure 3.5(b) which shows the benefit due to better-ear listening is greatly reduced compared to the anechoic case (Figure 3.2(b)). The benefit is limited to listening positions that are close to the interferer. Positions further away from the interferer have a lower direct-to-reverberant ratio for the interfering sound, and because reverberant sound tends to come from all directions the amount of energy arriving at both ears is more equal, reducing the ability to use better-ear listening to the same extent.

It is worth noting that the circular dip due to the “bright spot” is still evident, albeit

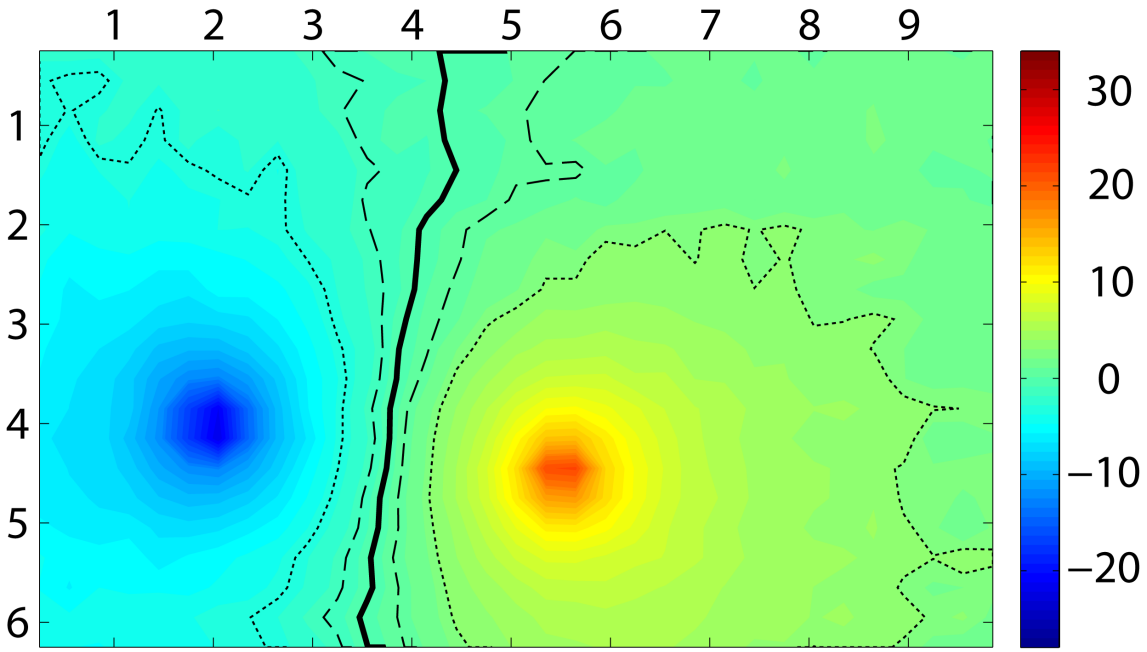
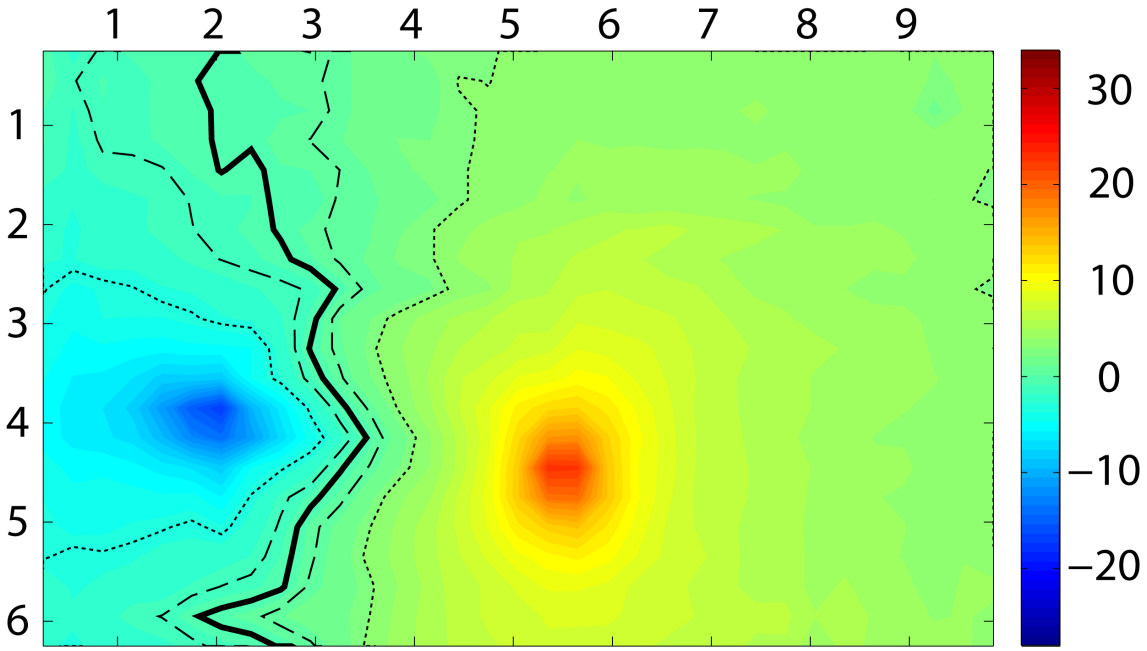
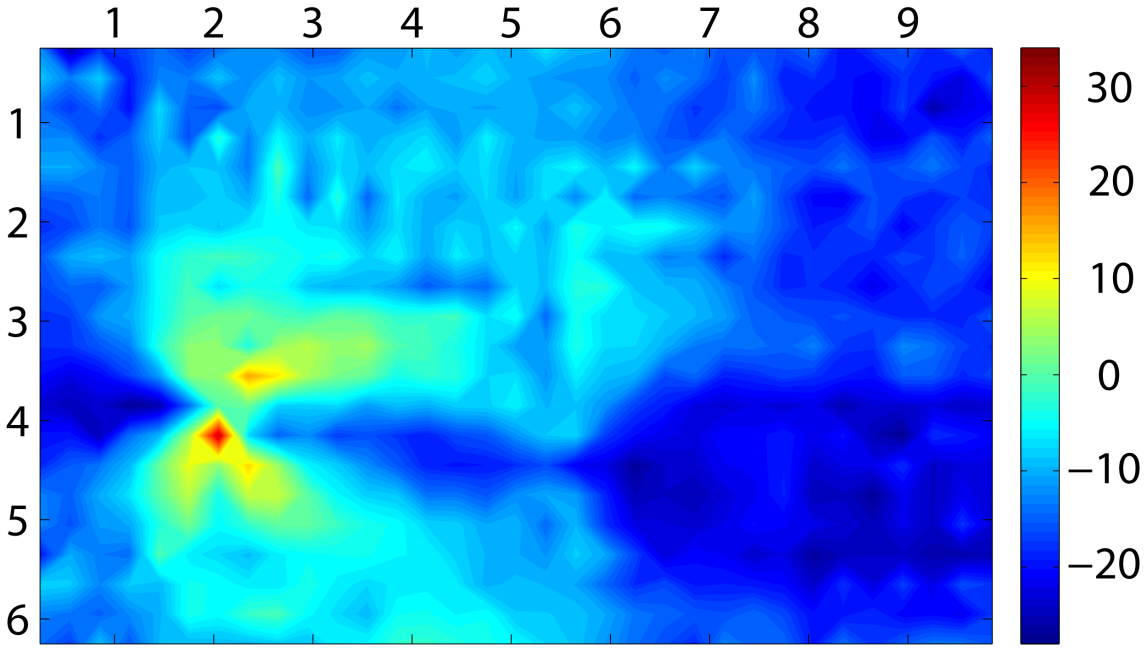


Figure 3.4: Map of TIR_{OMNI} in a reverberant 6.4m x 10m room, single interferer only, Target at 4.4 m by 5.5 m, Interferer at 3.9 m by 2 m.



(a) TIR_{BE}



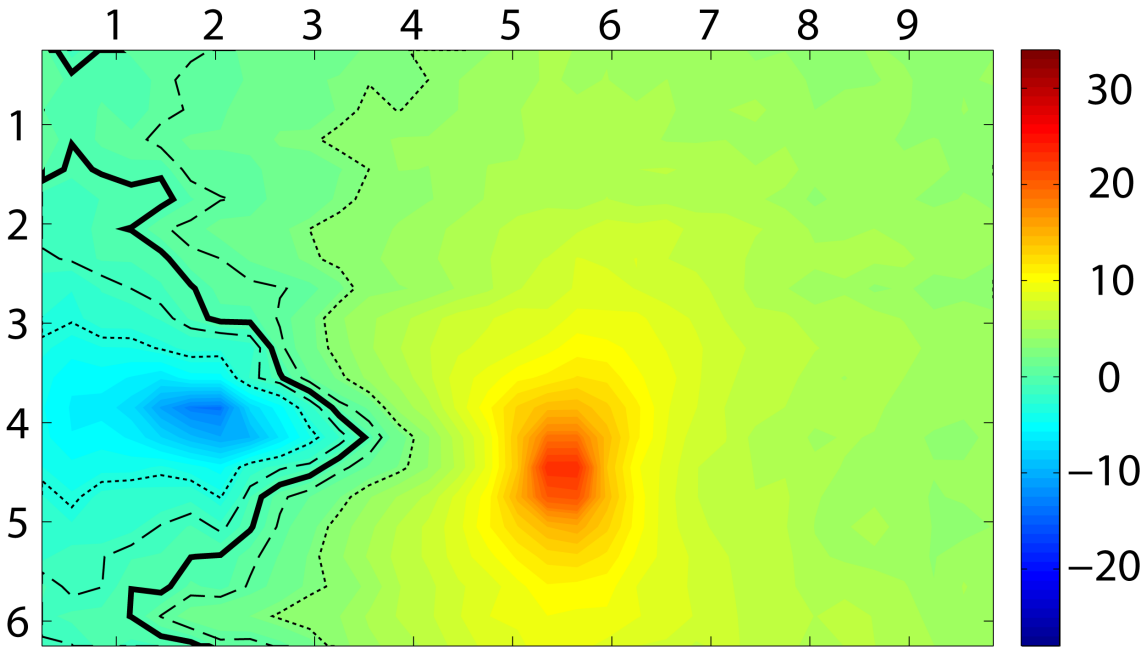
(b) B_{BE}

Figure 3.5: Map of (a) TIR_{BE} and (b) B_{BE} in a reverberant 6.4 m x 10 m room, single interferer only, Target at 4.4 m by 5.5 m, Interferer at 3.9 m by 2 m. The listener is always facing the target.

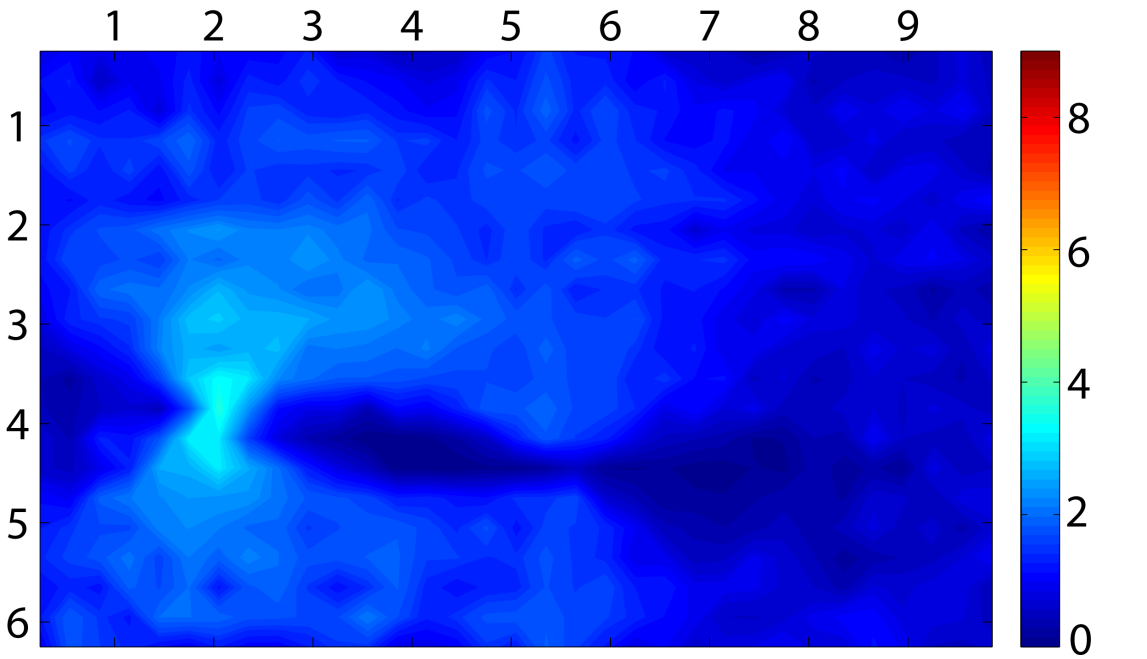
less obviously so. The benefit of better-ear listening is reduced in reverberation compared to the anechoic case; whilst the maximum value obtained is approximately the same, 8.4 dB in reverberation compared to 8.8 dB in the anechoic case, the mean is reduced, from 4.2 dB to 2.3 dB. Thus, for some listening positions within the room, reverberation is not detrimental to the ability to use better-ear listening, but the number of these locations is reduced.

3.1.2.3 Binaural Unmasking and Better-Ear Listening.

Figure 3.6(a) shows the results of the TIR_{BE+BU} modelling, and Figure 3.6(b) shows the B_{BU} (Fig. 3.6(a) minus Fig. 3.5(a)). The change between the results of the combined modelling and the better-ear-only modelling (Figure 3.5(a)) is small, pushing the 0 dB contour line closer to the interferer, but changing the $+3$ dB line only a small amount. This is evident in Figure 3.6(b), where, in comparison to the anechoic equivalent (Figure 3.3(b)), the level is greatly reduced, yet follows the same pattern in that it is most effective when the listening location is off the target-interferer axis. As with the better-ear listening effects, the benefit of binaural unmasking is greatest closest to the interferer. In this case, the high level of direct sound received by the listener results in a high interaural coherence, which facilitates binaural-unmasking. The level of binaural-unmasking available to the listener at the best location is comparable to that available in the anechoic case, 3.7 dB compared to 3.8 dB, but the overall amount available across the entire space is reduced, with a decrease in the mean from 2.7 dB to 1.2 dB with the introduction of reverberation.



(a) TIR_{BE+BU}



(b) B_{BU}

Figure 3.6: Map of (a) TIR_{BE+BU} and (b) B_{BU} in a reverberant 6.4 m x 10 m room, single interferer only, Target at 4.4 m by 5.5 m, Interferer at 3.9 m by 2 m. The listener is always facing the target.

3.2 Multiple-Interferer Configurations

A second configuration of sources was used to illustrate the capacity of the model to predict the effect of multiple interferers on spatial unmasking in anechoic and reverberant conditions. This second configuration had the same room specification as in the single interferer conditions, with the same size of room, absorption coefficients used on the wall surfaces, etc. The target was kept in the same position, as is the single interferer, but two further interferers are added at 1.4 m by 4 m and 2.9 m by 8.5 m. The three individual interfering sources and the target source had equal power levels. The design of this source configuration is such that it allows for comparison between single and multiple interferers, as well as the effects of reverberation.

3.2.1 Anechoic Space

The three-interferer anechoic configuration was simulated in the same manner as the single-interferer configuration, but with three interferer BRIR pairs concatenated together into a single longer BRIR pair.

3.2.1.1 Omnidirectional Maps

Figure 3.7 shows the results of modelling the TIR_{OMNI}. In the same way as with the single-interferer configuration, the dominant factor affecting the target-to-interferer ratios is the proximity of the listener to the three interferers, governed by the inverse relationship between distance and level. The effect of having three interferers compared to the single interferer is that the mean target-to-interferer ratio is reduced by 7.6 dB.

It is worth noting that each individual impulse response pair generated by the acoustical model has equal power; therefore each interferer has the same power as the target.

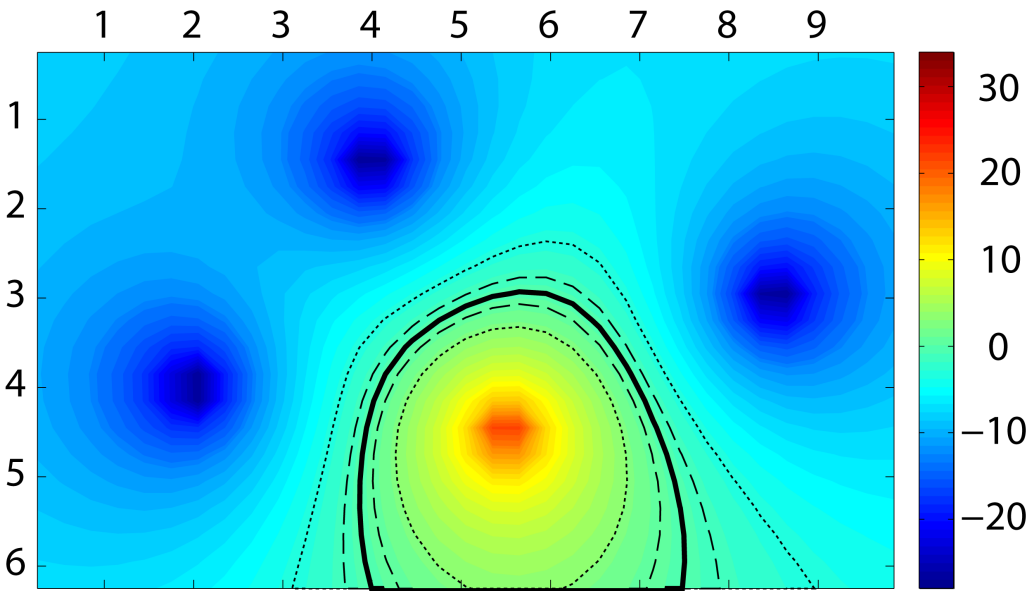
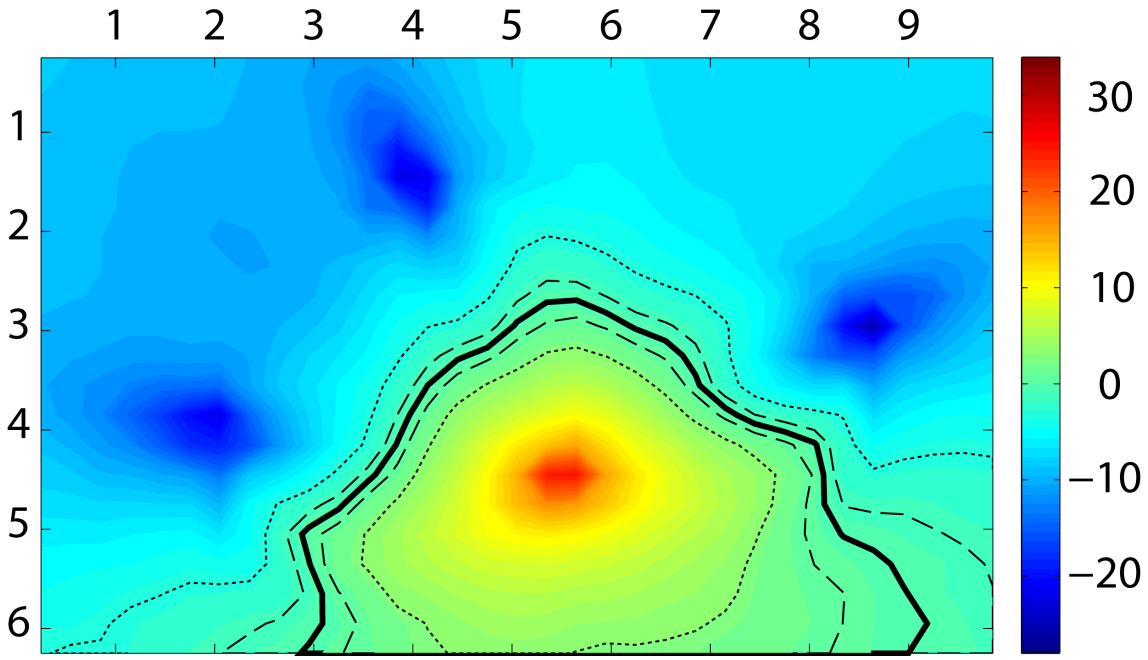


Figure 3.7: Map of TIR_{OMNI} in an anechoic 6.4 m x 10 m room, three interferers. Target at 4.4 m by 5.5 m, Interferers at 3.9 m by 2 m, 1.4 m by 4 m, and 2.9 m by 8.5 m.

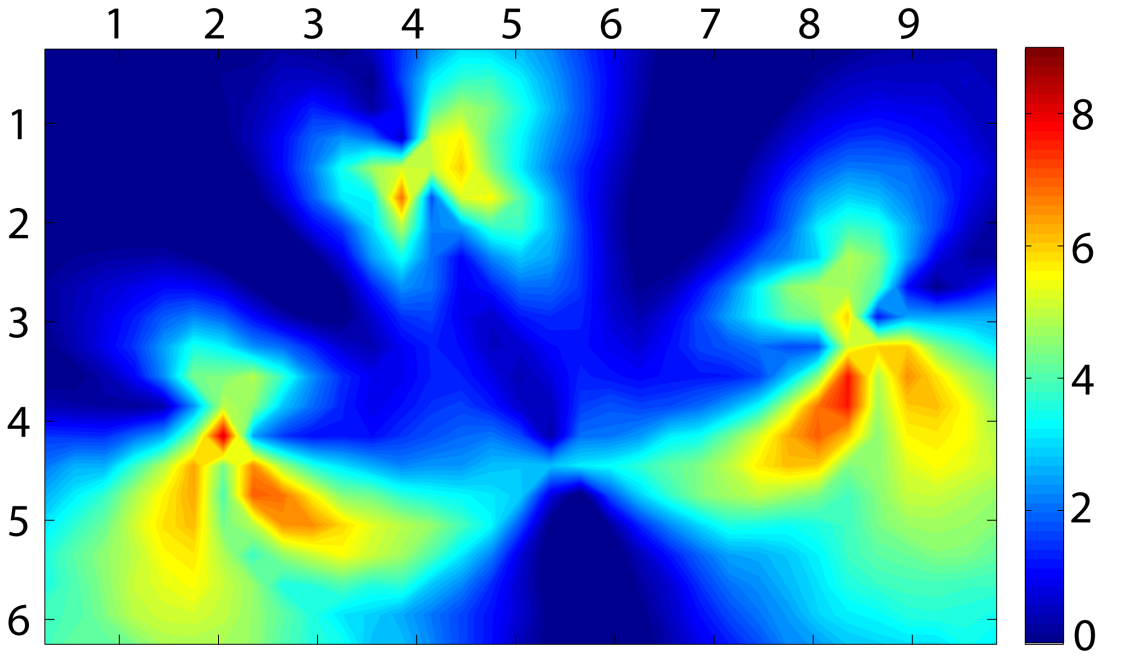
As such the concatenated interferer impulse responses, used for the calculations in the modelling, give an increase in total masking power of 4.77 dB compared to the single interferer conditions. Due to this it would be necessary to reduce the observed multiple interferer values by 4.77 dB for equivalent comparison (in the sense of equal total masking power) to the single interferer conditions. In this omnidirectional case, if the level of power from emitted from the three individual interferers was to be emitted from one source rather than three, then the reduction in mean target-to-interferer ratios can be corrected from 7.6 dB to 2.8 dB.

3.2.1.2 Better-Ear Listening Maps.

Figures 3.8(a) and 3.8(b) show the effective TIR_{BE} with three interferers, and the B_{BE}, respectively. From Figure 3.8(b) it can be seen that the benefit obtained is high in areas when the listener is close to one of the sources, be it target or interferer, and low on the axes between the target and any one of the interferers. This is due to the fact that in areas that are remote from any one of the interferers the levels of the three interferers at the listener's head will be approximately equal, and often from differing directions, meaning that whilst the head may shadow one ear from one interferer this will have little impact on the overall TIR, whereas when the listener is near to an interfering source the effect of head shadow on that source will have greater impact. The mean benefit due to better-ear listening (B_{BE}) available in anechoic conditions is reduced from 4.2 dB to 2.2 dB by the addition of two further interferers. This benefit is not affected by the increase in the power contained within the concatenated impulse responses as it represents the difference in effective TIR between listening with two ears and listening through an omnidirectional microphone, both of which will show reduced TIRs following the addition of further



(a) TIR_{BE}



(b) B_{BE}

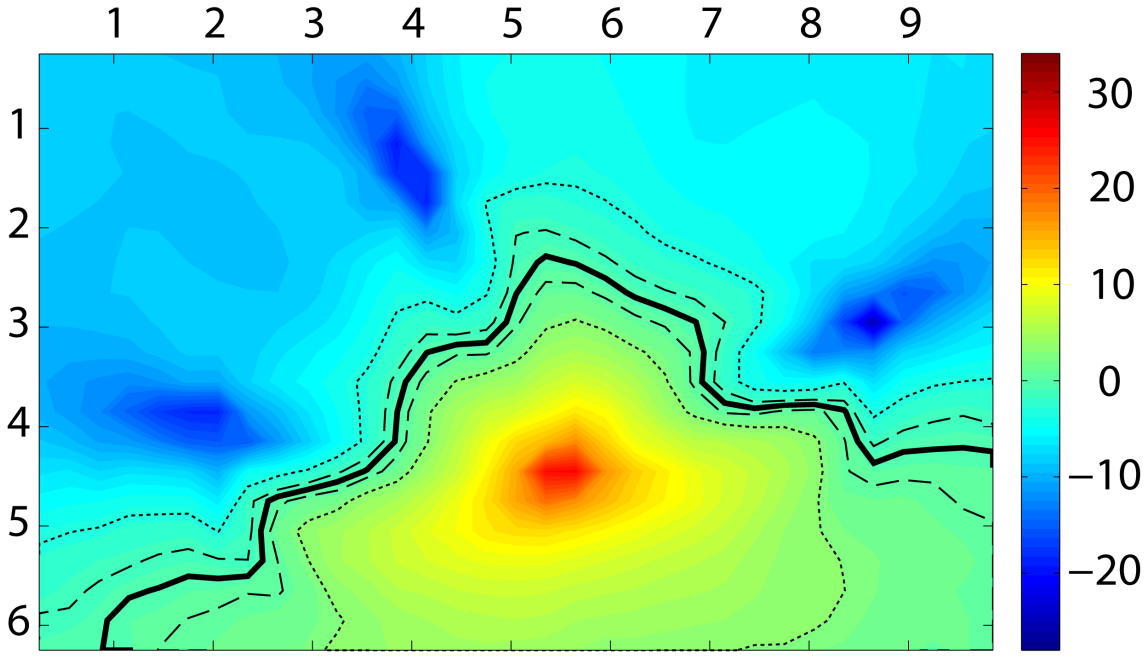
Figure 3.8: Map of (a) TIR_{BE} and (c) B_{BE} in an anechoic 6.4 m x 10 m room, three interferers, Target at 4.4 m by 5.5 m, Interferers at 3.9 m by 2 m, 1.4 m by 4 m, and 2.9 m by 8.5 m. The listener is always facing the target.

interferers.

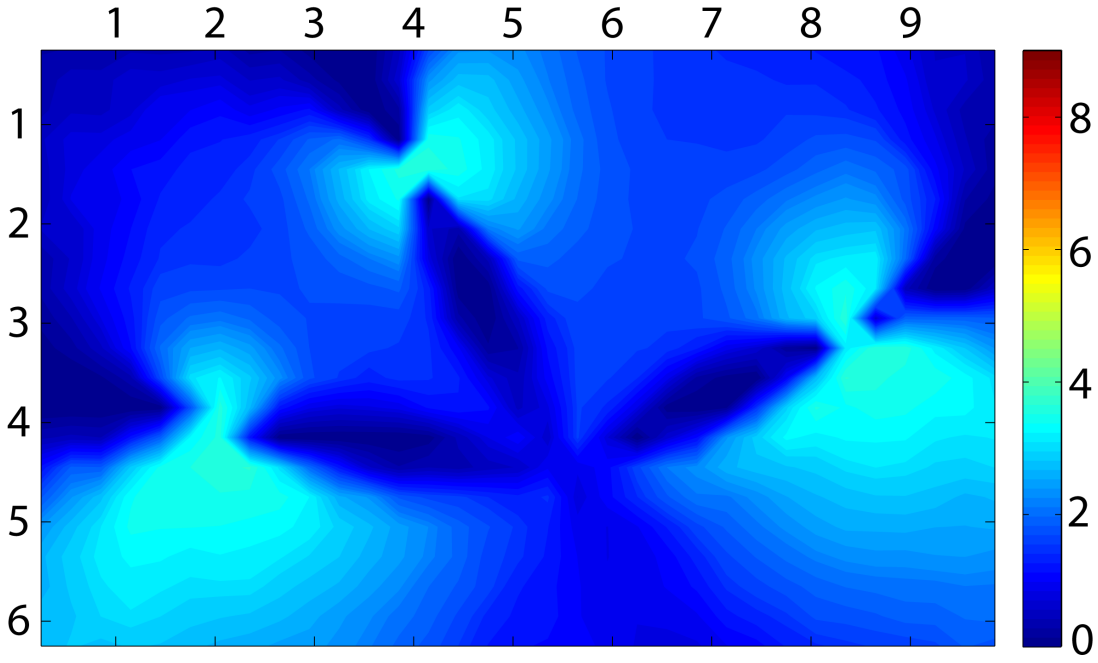
3.2.1.3 Binaural Unmasking and Better-Ear Listening.

Figures 3.9(a) and 3.9(b) show the results of the combined modelling of the TIR_{BE+BU}, and the B_{BU} respectively. The principal effect of binaural unmasking is to push the 0 dB contour line in figure 3.9(a) closer again to the interferers, and to increase the areas within which the level of the TIR is greater than 0 dB.

Figure 3.9(b) shows the pattern in which binaural-unmasking is effective across the space (Fig 3.9(a) minus 3.8(a)). As with the B_{BE}, the area in which binaural-unmasking is effective is reduced to those areas that are close to the interferers, and off the target-interferer axes. This is due to those areas having high levels of interaural coherence for the interferer to which they are close. At the greater distances the interaural coherence is reduced because three sets of signals arrive at the listener from differing angles, with approximately equal power. The maximal level of binaural-unmasking available to the listener is high, 3.7 dB, but the mean level is reduced by the inclusion of two more interferers, reduced to 1.7 dB compared to 2.7 dB in the single interferer condition.



(a) TIR_{BE+BU}



(b) B_{BU}

Figure 3.9: Map of (a) TIR_{BE+BU} and (b) B_{BU} in an anechoic 6.4 m x 10 m room, three interferers, Target at 4.4 m by 5.5 m, Interferers at 3.9 m by 2 m, 1.4 m by 4 m, and 2.9 m by 8.5 m. The listener is always facing the target.

3.2.2 Reverberant Space

The modelling of the multiple-interferer configuration in a reverberant space followed the same method as that used for the single interferer configuration, allowing for a comparison between single and multiple interferers, as well as between multiple anechoic and reverberant interferers.

3.2.2.1 Omnidirectional Maps

Figure 3.10 shows the results of modelling the omnidirectional listening condition for a single microphone in space. Compared to the anechoic condition (Figure 3.7) the 0 dB contour line is closer to the target, whilst the distances between the $\pm$1 dB and $\pm$3 dB lines are similar. This shows that the overall level of TIR has been reduced but the rate of change with distance from the target is approximately constant.

3.2.2.2 Better-Ear Listening Only.

Figures 3.11(a) and 3.11(b) show the effective TIR_{BE} and the benefit of better-ear listening over the omnidirectional condition (Fig. 3.11(a) minus Fig. 3.10). The overall pattern of benefit is the same as that observed in the anechoic multiple-interferer condition (Fig. 3.8(b)), but reduced in degree. The mean TIR_{BE} is reduced by 1 dB.

3.2.2.3 Binaural Unmasking and Better-Ear Listening.

Figures 3.12(a) and 3.12(b) show the result of modelling the combined TIR_{BE+BU} effects, and the B_{BU} respectively. The effect of binaural unmasking is to improve the effective TIR by a mean of 1.1 dB compared to better-ear listening alone. The pattern of benefit follows that of the anechoic multiple interferer configuration, but is reduced

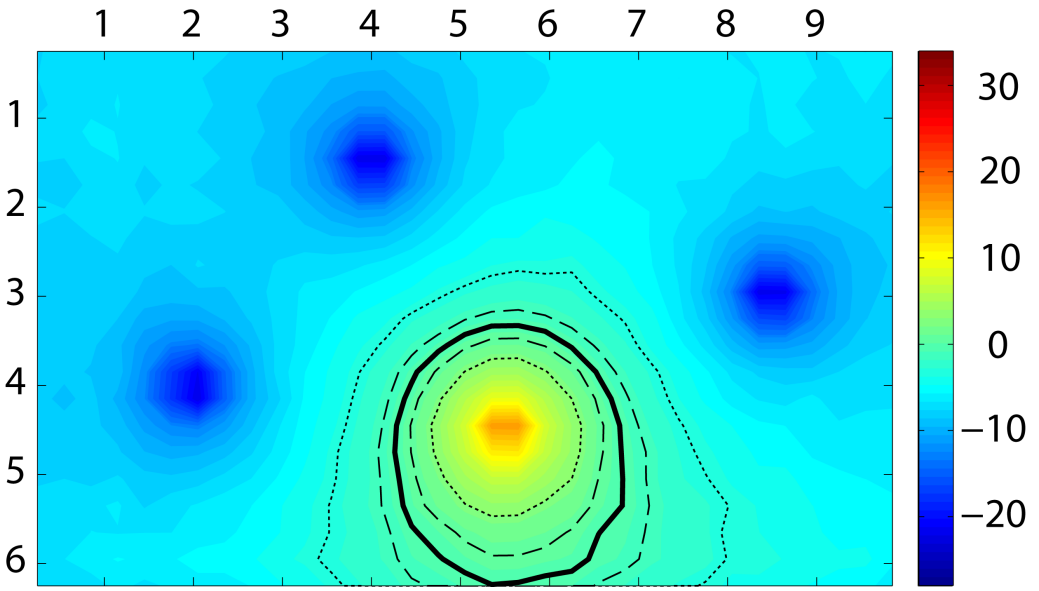
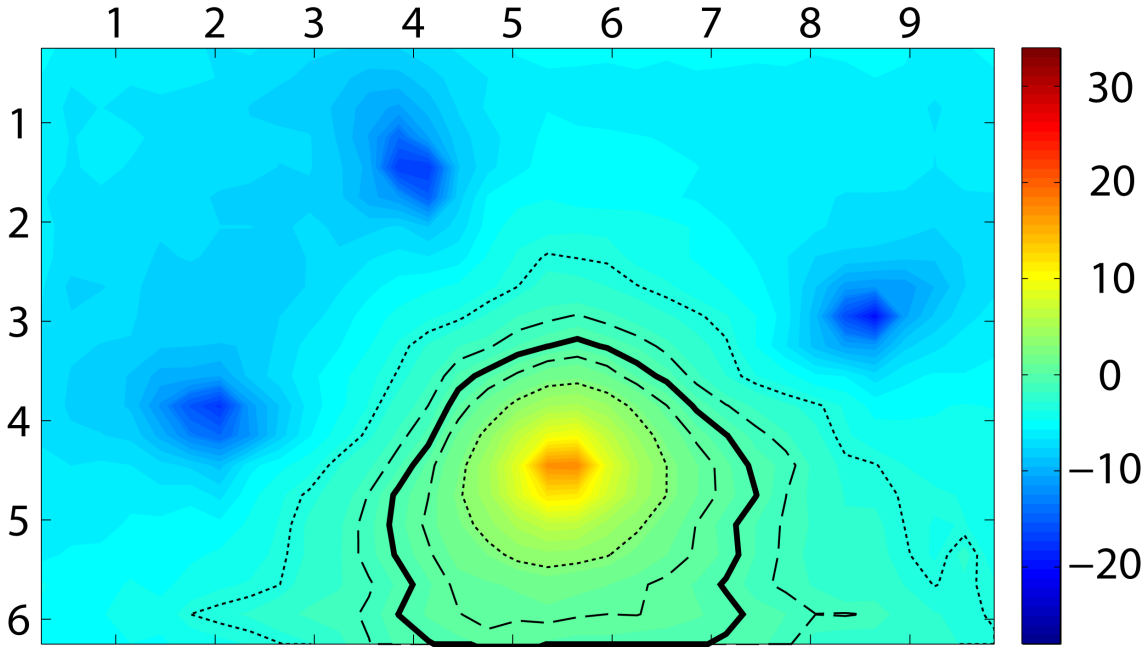
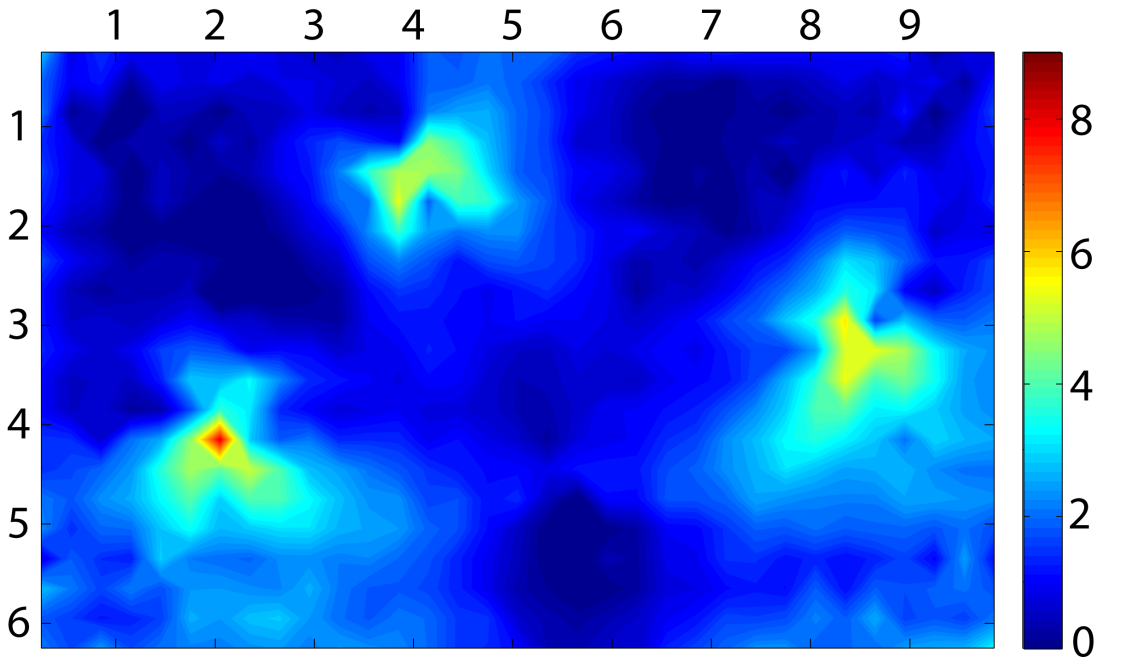


Figure 3.10: Map of $\mathrm{TIR}_{\mathrm{OMNI}}$ in a reverberant 6.4 m x 10 m room, three interferers. Target at 4.4 m by 5.5 m, Interferers at 3.9 m by 2 m, 1.4 m by 4 m, and 2.9 m by 8.5 m.

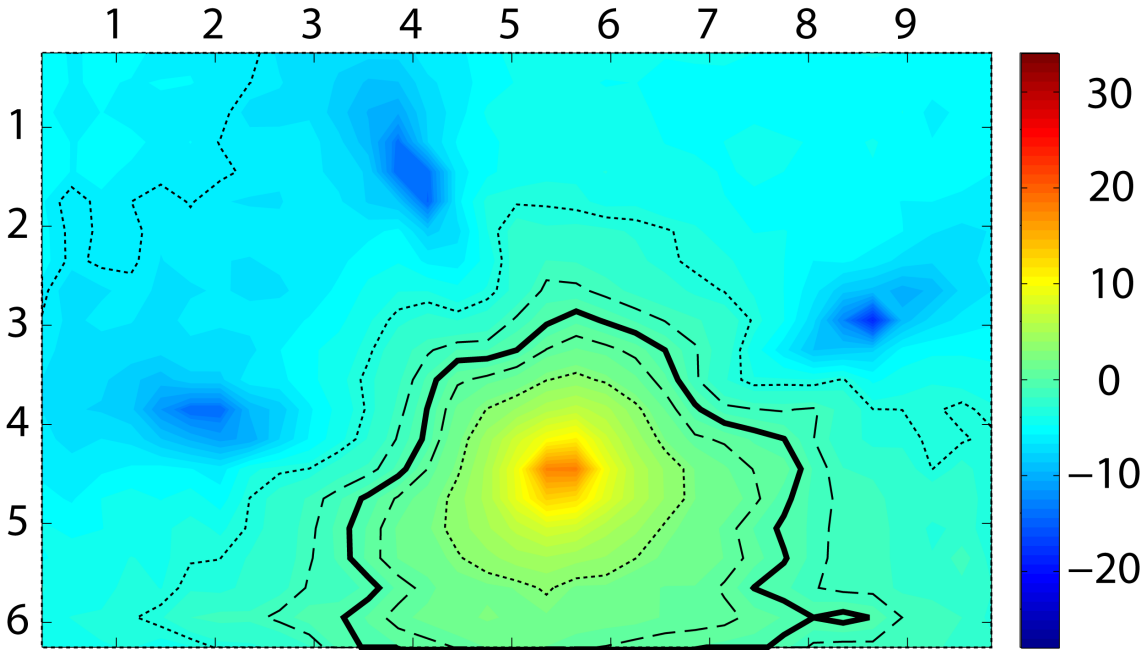


(a) TIR_{BE}

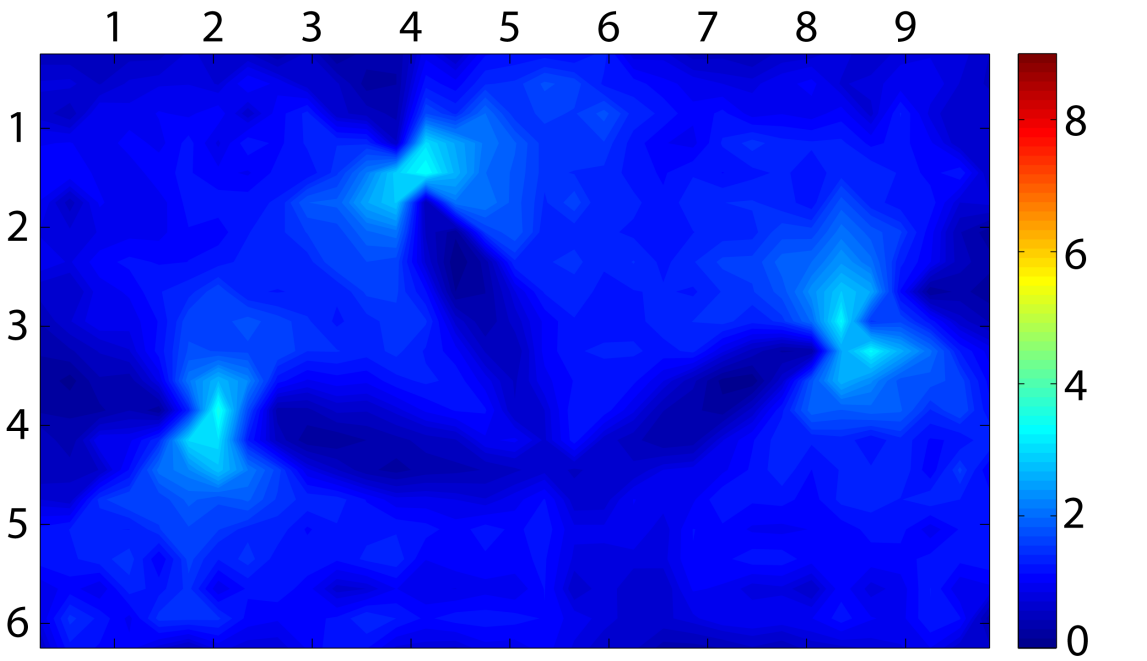


(b) B_{BE}

Figure 3.11: Map of (a) TIR_{BE} and (b) B_{BE} in a reverberant 6.4 m x 10 m room, three interferers, Target at 4.4 m by 5.5 m, Interferers at 3.9 m by 2 m, 1.4 m by 4 m, and 2.9 m by 8.5 m. The listener is always facing the target.



(a) TIR_{BE+BU}



(b) B_{BU}

Figure 3.12: Map of (a) TIR_{BE+BU} and (b) B_{BU} in a reverberant 6.4 m x 10 m room, three interferers, Target at 4.4 m by 5.5 m, Interferers at 3.9 m by 2 m, 1.4 m by 4 m, and 2.9 m by 8.5 m. The listener is always facing the target.

in degree. The maximum benefit of binaural-unmasking available is reduced by 0.2 dB compared to the anechoic case, whilst the mean level is reduced by 0.6 dB.

3.3 Discussion

Two questions were posed at the opening of this chapter; what is the effect of reverberation on the listeners ability to separate competing sources, and what is the effect of multiple interferers on that ability?

3.3.1 The Effect of Reverberation

The omnidirectional maps show us the simple effect of reverberation on the sound levels within the space. For the single interferer conditions reverberation reduces the mean predicted target-to-interferer ratio, as well as reducing the standard deviation of the target-to-interferer ratio (see Table A.4, p. 189). On average the listener will not be able to perform to the same level as the anechoic space in the reverberant space. For areas where they are close to the target the listener will perform worse in reverberation than in the anechoic case, whilst for areas close to the interferer their performance will increase. For the multiple interferer conditions an *increase* in the mean target-to-interferer ratio is observed, as well as a decrease in the accompanying standard deviation, showing that the overall effect of reverberation is of benefit to the listener, at least for a listener with neither the benefit of better-ear listening or binaural unmasking. One caveat to this is that all of the energy in the target IR is assumed to be of benefit to the listener, which is unlikely to be true. Bradley (1986*a*, 1986*b*) used the Useful-to-Detrimental ratio (U) to predict the intrinsic effect of reverberation on the intelligibility of a target speech source, showing that the later arriving energy of the signal was detrimental to understanding. At short

distances between target and listener the U will be high, and the accuracy of the model will be good, but at greater source distances the U will decrease, and the accuracy of the model will also likely decrease.

The benefit due to better-ear listening in the single-interferer conditions is reduced in both mean and standard deviation by the addition of reverberation, but the maximum observed benefit remains approximately the same (Table A.5, p. 189). Though the direct sound will be unaffected by the addition of reverberation, the smoothing of the sound field means that the difference in source levels at the listeners left and right ears when distant to any one source is reduced, limiting the ability of better-ear listening to reduce the level of the interferer in any situations other than when the listener is near to one of the sources, be it target or interferer. Because of this the maximum observed benefit of better-ear listening remains approximately constant, but the area in which a listener can make the most of this is reduced. The areas where maximal benefit due to better-ear listening occurs are those close to the interferer, but then these are also the areas where it is most required. Better-ear listening thus helps most in the more challenging circumstances.

The benefit of binaural unmasking follows the same pattern as the benefit of better-ear listening, in that the addition of reverberation to either the single or multiple interferer conditions reduced the mean and standard deviation of the predicted benefits, whilst not affecting the maximum benefits. Again, the areas in which binaural-unmasking is greatest are those in which the worst performing target-to-interferer ratios occurs.

Overall, the effect of reverberation in the single interferer configurations is to reduce the modelled $\mathrm{TIR}_{\mathrm{BE+BU}}$, through both reducing the $\mathrm{TIR}_{\mathrm{OMNI}}$, as well as reducing the benefit of better-ear listening and binaural unmasking. Whilst in the multiple interferer configurations the effect of reverberation is to increase the modelled $\mathrm{TIR}_{\mathrm{OMNI}}$, compared to

the anechoic equivalent, the reduction in benefit due to better-ear listening and binaural unmasking means that the overall TIR_{BE+BU} is reduced.

3.3.2 The Effect of Multiple Interferers.

The omnidirectional maps show us the simple effect of adding multiple interferers to the model. Comparing the anechoic single interferer condition to the anechoic multiple interferer condition there is a reduction in the mean target-to-interferer ratio. Combined with a reduction of the standard deviation this can be attributed to the fact that there is less area in which the listener is close to the target, yet distant from all of the interferers. In reverberation, the addition of extra interferers reduces the mean target-to-interferer ratio, but there is little effect on the standard deviation. Given that reverberation can be viewed as many hundreds of relatively distant anechoic sources, (one source per reflection), adding extra interferers to a single reverberant interferer will increase the number of these distant sources, moving from hundreds to possibly thousands of sources. However, this will have less of an impact than the original addition of reverberation to the single interferer, where it moved from only the single interfering source to many hundreds of sources.

For the benefits due to both better-ear listening and binaural unmasking, the addition of multiple interferers reduces the observed mean levels when comparing anechoic to anechoic, and reverberant to reverberant, but in all cases the maximum observed benefit remains constant. The addition of multiple interferers to the anechoic case has the same impact on the mean benefit of better-ear listening and binaural unmasking as the addition of reverberation. This outcome is consistent with the concept of reverberation as simply the addition of multiple anechoic interferers. The addition of reverberation to a single

interferer, is the same as adding a number of anechoic interferers. One note though, the model does not take account of the temporal window of the auditory system, and assumes that all energy in the target is useful, when in reality a certain amount of the energy in a reverberant target will de detrimental to the ability of a listener, the amount of which will vary with C_{50} / RT_{60}.

As with the effect of adding reverberation, the addition of multiple interferers reduces the areas in which large benefits are obtainable, but not the maximum benefit that can be obtained. Both better-ear listening and binaural-unmasking are still available in multiple interferer conditions, but they are reduced to areas which are close to one of the interfering sound sources. Free movement within the space would allow the listener to obtain greater target-to-interferer ratios in most areas than if they were to remain stationary, but where free movement is not possible, and particularly in the worst performing areas, better-ear listening effects and binaural-unmasking are still of use to the listener, albeit at a reduced level.

3.3.3 Limitations

The modelling carried out in this chapter tells us about the effects of multiple interferers and of reverberation on the ability of a listener to use spatial hearing to listen to a target sound in the presence of an interfering noise source. However, there are a number of other questions that could be answered.

The architecture of the room will play a large role in determining how detrimental reverberation will be on the listener's performance. With a larger room, and therefore a longer reverberation time, reverberation is likely to have a greater effect, smoothing the difference in sound levels between the left and right ears further. The geometry of the

room will also play a key role, especially if concave curved walls, or a domed ceiling were to be used, focussing the reflected sound into particularly good areas (when focussing the target) or particularly poor areas (when focussing the interferer(s)).

Whilst the present modelling shows the difference in effects between either one or three interferers, the findings may not hold true when there is a greater number of interferers. Bronkhorst and Plomp (1992) show large differences in anechoic speech reception thresholds for having either one or two interferers, but with increasing numbers of interferers the relative effect decreases. Beyond a certain point the number of interferers will be so great as to create the effect of a single diffuse source. At this stage the addition of further interferers will have no effect, beyond increasing the level of energy contained within the interferer.

Within the model the listener was always positioned so as to be facing the target, whilst both better-ear listening and binaural unmasking require a spatial separation of the target and the interferer. Therefore, in order to maximise the benefits to the listener, particularly of better-ear listening, it may be beneficial that the listener does not face the target. With free rotation of the head the listener may be able to maximise the effects of head-shadow on the interferer, creating a larger interaural level difference, and so making greater use of better-ear listening. At the same time this free rotation may allow the listener to create larger interaural phase differences for the target and the interferer, particularly when the sources are located with one directly in front of them, and one directly behind them, allowing the mechanisms predicted by equalization-cancellation theory to give higher target-to-interferer ratios. When free movement within the space is not possible, head rotations may allow the listener to improve their ability to understand the target speech. The constraint on head rotation will be from both the social implications of this, it is

generally accepted that you face the person whom you are listening to, and the loss of visual cues that aid understanding, such as lip reading.

Chapter 4

Head Orientation

In the previous sections, all modelling of target-to-interferer ratios was done with the listener facing the target, but as discussed in Chapter 3, allowing listeners to freely rotate their head may improve their ability to understand speech in noise. Listeners can then find the optimum head orientation for which they can make use of better-ear listening and binaural unmasking to help them best understand the target speech.

In this chapter the effect of head rotation on the benefit of better-ear listening and binaural unmasking is examined in two ways. Firstly, the predicted benefits with respect to listener orientation are modelled in a specific set of target and interferer conditions, while systematically varying the relative azimuthal position of the target, the interferer, and the number of interferers. Secondly, a number of spatial maps similar to those used in chapter 3 are modelled, this time to include the effects of head rotation by allowing the listener to optimise head orientation for each position in space.

4.1 Polar Plots

To examine the way in which the orientation of a listener to anechoic target and interferer sources affects the listeners' ability to understand speech in noise, a number of polar plots were generated. This was done using the HRTFs of a KEMAR manikin (as recorded by researchers at MIT (Gardner & Martin 1994)) as the input to the revised model, using the same method as used in Chapter 2 to model the data of Peissig and Kollmeier (1997) (See 2.2.2, p. 33). By varying which HRTFs were used for the target and the interferer(s), their respective azimuth locations were changed, rotating them around the listener.

4.1.1 Effect of Azimuthal Separation

To begin investigating the effects of head orientation on listener performance, polar patterns of benefit were first produced with a fixed listener, whilst varying the azimuth of one or both sources through a full 360° range. This approach is used to gain insight into the differing effects of individual target and interferer azimuths alone, before looking at the effect of the combined target and interferer azimuth.

The first polar plots show the effects of having (1) the target in front of the listener and varying the interferer azimuth, (2) the interferer located in front and varying the target azimuth, or (3) having them at equal and opposite azimuths ("symmetrical"). These effects are shown in figure 4.1. The plots show that it is of greater benefit to have a target in front and a spatially separated interferer than the interferer in front and a spatially separated target. The red line is the predicted B_{BE+BU} for a target source fixed at 0°, directly in front of the listener, with an interfering source rotated around the listener. The value of the line on the radial axis is the B_{BE+BU} for when the interferer is located at that angle. The blue line is the same as the red line, except that in this instance the interferer is fixed

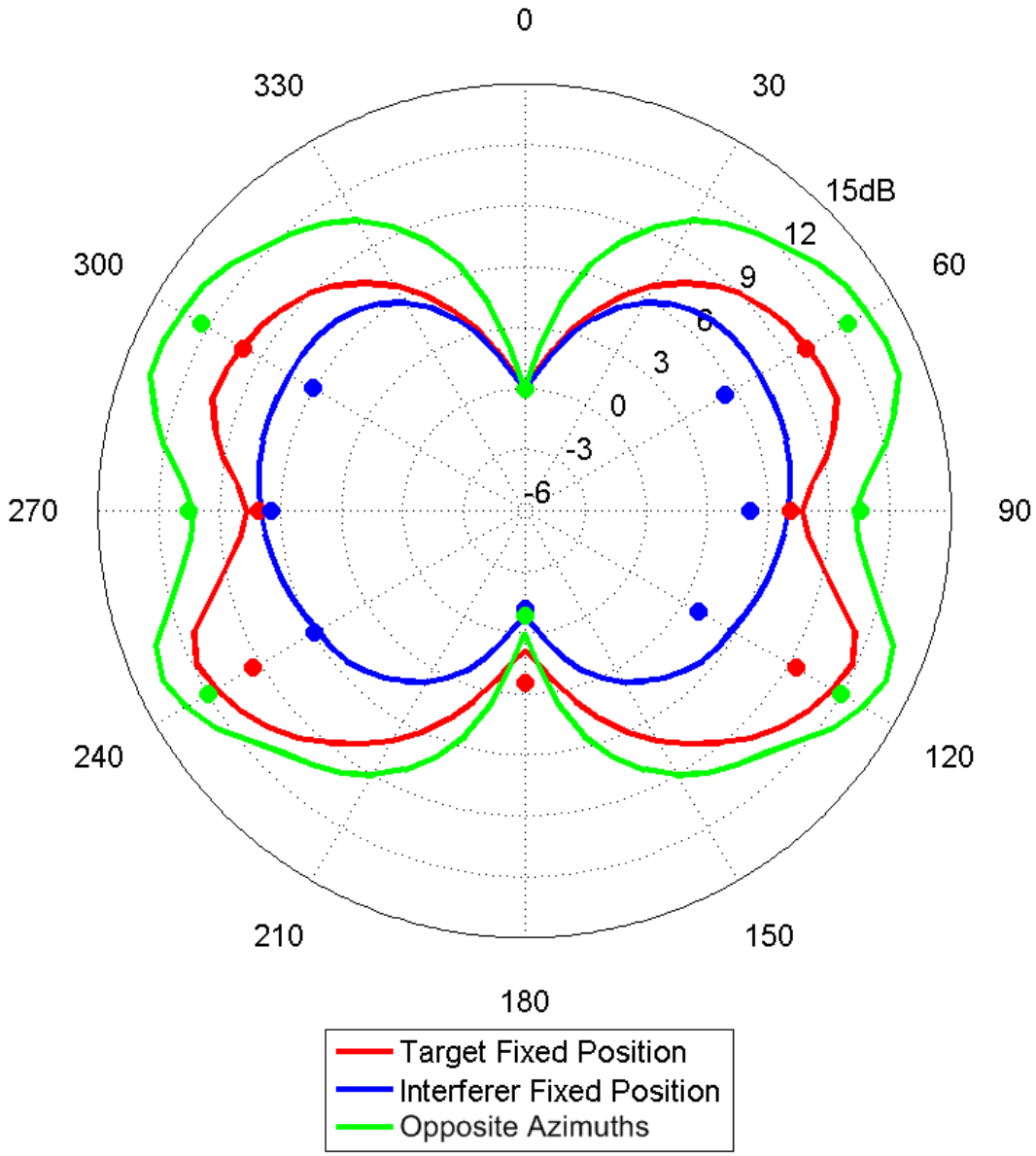


Figure 4.1: Polar plots of B_{BE+BU} with respect to source azimuth, either having a target fixed at $0°$ and moving interferer, an interferer fixed at $0°$ and moving target, or with neither source fixed, both moving around the listener in equal and opposite directions. Filled data points are from empirical data supplied by Juan-Pablo Ramirez of Deutsche Telekom Laboratories, Berlin Institute of Technology. r = 0.98, $p < 0.0001$

at 0° while the target is rotated about the listener. For the final green line the target and interferer have equal and opposite azimuths so that, for example, the value of the line at 30° is for a target at 30° to the listener, and an interferer at 330°. The azimuthal resolution of the data is in 5° increments (the resolution of the measured HRTF plane data sets), and only the horizontal plane data was used; no effects of elevation are included.

The filled circular data points are from empirical data collected by Juan-Pablo Ramirez of Deutsche Telekom Laboratories (Berlin Institute of Technology), who measured SRTs for short semantically unpredicted sentences presented with a speech-shaped noise interferer. The correlation coefficient between the measured data points used and the equivalent modelled data points is 0.98 ($p < 0.0001$).

Looking first at the target-in-front line (red), at the 0° position there is no B_{BE+BU} due to the fact that the target and the interfering source are collocated, and therefore there is no benefit due to better-ear listening or binaural unmasking. As the interferer is moved away from the target, the modelled B_{BE+BU} increases rapidly, to a peak of 10.3 dB at 70°, before decreasing to 7.6 dB at 90°. This reduction in B_{BE+BU} at 90° is due to the effects of constructive interference at the contralateral ear, as shown in the previous modelling of the data sets from the literature (see the Discussion section in Chapter 2, p. 42). As the interfering source moves behind the listener the predicted B_{BE+BU} increases to 11.8 dB at 115°, which is 1.5 dB higher than the maximum when the interferer is in front of the listener. This is due to a greater B_{BE} behind the listener, possibly caused by the contribution of the pinnae to front-back level differences. This is also evident when the interferer is directly behind the listener (180°), as the B_{BE+BU} at this point is still 0.8 dB. With the interferer behind the listener there will be no effects of head shadow on the interferer, and as such no ILD for either the target or the interferer. There is also no B_{BU},

because there is no ITD for either the target or the interferer sources. Consequently, the positive B_{BE+BU} must be caused by the additional filtering of the interfering source caused by the pinnae as included in the HRTFs. The left and right sides of the polar plot are mirror images of each other, this perfect symmetry is unlikely to hold true in real world situations where real human HRTFs are used, and is a factor of using symmetrical HRTFs measured from a symmetrical dummy head. That said, any deviations from true symmetry are likely to be small compared to the azimuthal variations.

With the interferer fixed in front of the listener and the target rotated about the listener (the blue line) a very different pattern of B_{BE+BU} is created, being both reduced in level compared to the target-in-front condition, and lacking the dip at $\pm 90^{\circ}$. The B_{BU} that is produced by having the interferer fixed in front of the listener will be much the same as that for having the target fixed in front. Because the differences in interaural phase created by moving a source are similar for both the target or the interferer, the ability to use binaural unmasking remains the same. The difference in B_{BE+BU} between the interferer-in-front and the target-in-front conditions is thus due to the difference in B_{BE}. With the target in front and the interferer to one side of the listener, a higher B_{BE} will occur at the ear occluded from the interferer, whereas with the interferer in front and the target to one side neither ear is occluded from the interferer, so the listener can only use the ear which receives the most energy from the target.

For the symmetrical situation, both target and interferer are rotated about the listener in opposite directions (the green line). Here, a pattern similar to that of the fixed target condition is produced, but with larger effects and a maximum value of 13.7 dB. Again, the B_{BU} is similar to that which occurred in the other two conditions, and the changes in B_{BE+BU} are created by the B_{BE}. As the sources move apart, the ear that is occluded

from the interferer is at the same time pointed more directly towards the target, and so the B_{BE+BU} increases to a greater level than when the target is in front.

For a listener to understand speech in a noisy environment there is greater levels of benefit to be gained with either a target directly in front of them and the interferer off to one side, or with both sources distributed to either side, than there is to be gained by the listener facing the interferer and having the target situated elsewhere. But, if the azimuthal separation of the sources is fixed, at say $90°$, is it better to have the target in front and the interferer to one side, or to face neither and have the target at, for example, $30°$and the interferer at $120°$?

4.1.2 Hawley *et al* Configurations

The data of Hawley *et al* (2004) has already been predicted accurately (r = 0.99, $p < 0.0001$) with the revised methodology (See Chapter 2.2.3, p. 35). The same configuration of sources will be used in the following modelling, using either one, two, or three interferers, but now rotating the listener within this configuration; the relative azimuthal angles of the sources remain constant. This allows for investigation of the effects of listener orientation, within a fixed field of sources, such as when they are positioned within a room and a listener is attempting to listen to a talker in the presence of a noise interferer, but is unable to move closer to the target. Unlike the multiple interferer modelling carried out in chapter 3, the combined power of the interferers is controlled so as to be equal to that of the single target, irrespective of the number of sources. The reason for this is so as to make the comparison between different numbers of interferers easier, a single interferer, or three collocated interferers will produce the same response.

4.1.2.1 One interferer

Figure 4.2 shows the results of modelling the effect of head orientation in Hawley *et al*'s (2004) single interferer conditions, with a noise interferer. In each case the target source is located at $0°$. Panel (a) shows the condition when the interferer is located at $0°$, (b) is with the interferer at $330°$, (c) is with the interferer at $60°$, and (d) is with the interferer at $90°$. The lines show the overall B_{BE+BU}, the B_{BE}, and the B_{BU}. If you take the B_{BE+BU} data point at $0°$ from each plot they give the same results as those predicted in chapter 2.2.3, except expressed as benefit values rather than SRTs, which are equal and opposite to each other.

As expected, for the case when both the interferer and the target are collocated at $0°$ (fig 4.2(a)), there is no effect of head orientation, with the B_{BE+BU} consistently 0 dB, due to their being no difference between target and interferer in ITD or ILD at any angle.

With the interferer located at $330°$ (fig. 4.2(b)), there is an effect of head orientation. When the listener faces the target a B_{BE+BU} of 6.77 dB is predicted, with 3 dB of it due to the B_{BU}. As the listener rotates clockwise the B_{BE+BU} remains constant until approximately $30°$, where it starts to gently decrease to approximately 1.0 dB at the $75°$ point. This reduction is caused by a reduction in the both the B_{BE} and B_{BU}. This is because at this point both sources are located to the same side of the listener, one at $75°$ to the listener, and the other at $105°$. At this point, they produce equal ITDs and approximately equal ILDs as they are equal amounts forward or back from the listener, this results in no B_{BU}, and there is only a small B_{BE} due to the front-back differences in the pinnae. As the listener rotates past the $75°$ position, the B_{BE+BU} rapidly increases as the target reaches $90°$ to the listener, due to a high B_{BE}. The B_{BE+BU} remains high as the listener rotates until $210°$, where it decreases rapidly to zero dB at $240°$, where again the target and inter-

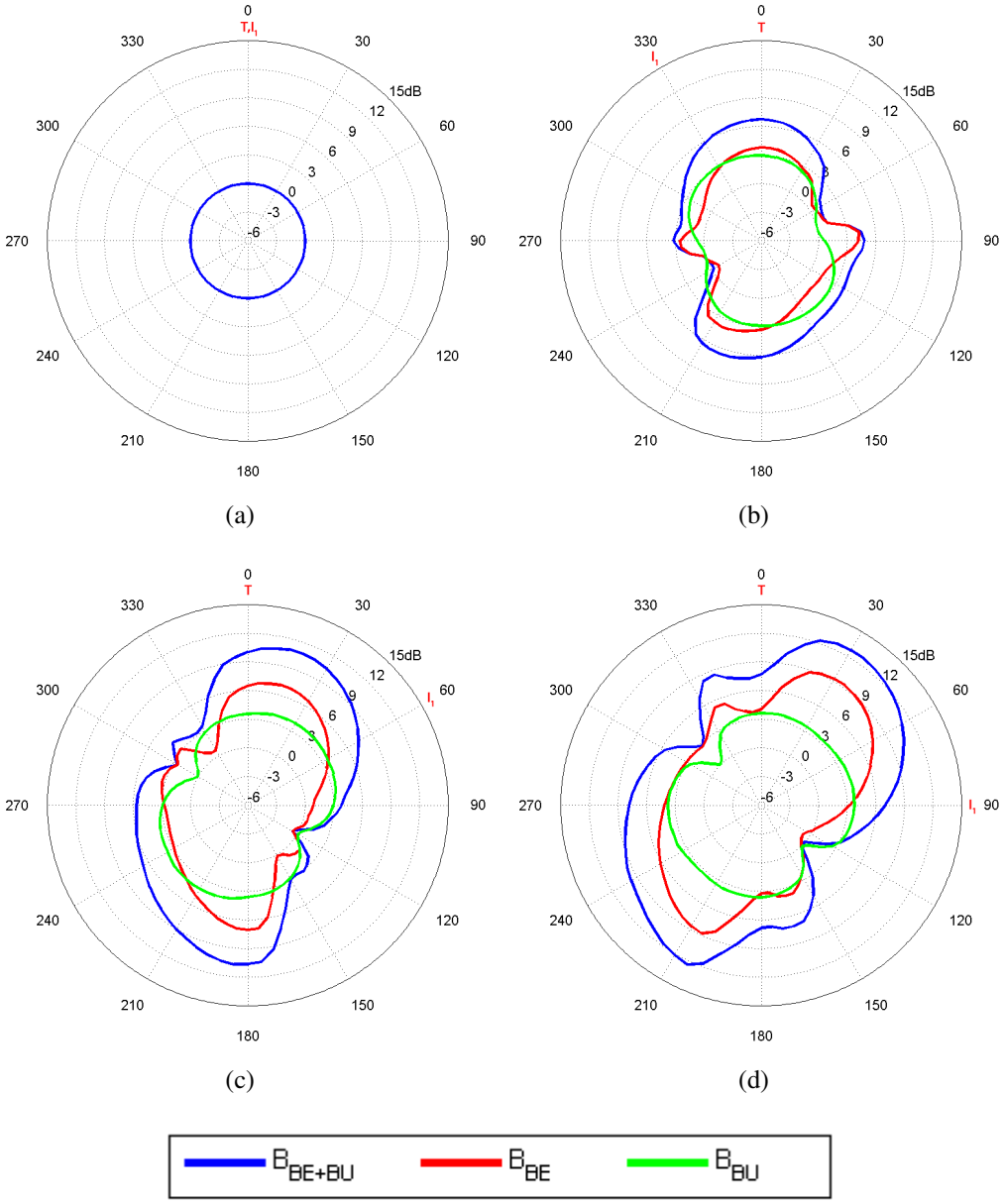


Figure 4.2: Head orientation polar plots, Hawley *et al* (2004) single interferer conditions, with target at 0° and interferers at: (a) 0°, (b) 330°, (c) 60°, (d) 90°

ferer lie equally forward and back from the interaural axis, reducing the B_{BE} and the B_{BU}. As soon as the listener rotates further the TIR_{BE+BU} recovers, and remains high until the listener completes the rotation.

Figure 4.2(c) shows the effects of listener rotation when a single interferer is present at a fixed position of $60°$. The overall pattern of B_{BE+BU} and benefits is the same as that produced in the $330°$ interferer condition (Fig. 4.2(b)), though with an increased maximum level, and rotated so that the peak of the B_{BE+BU} is at approximately $15°$, and the nulls in the pattern of the B_{BU} occur at $120°$ and $300°$. As with the previous plot this is due to this axis being the position where the target and the interferer are located on the same side of the listener, and symmetrically placed about $90°$.

As at $60°$, when the interferer is fixed at $90°$ the pattern produced by modelling the B_{BE+BU} is the same as for the $330°$ configuration, but again rotated further clockwise, towards the interferer, and with a greater maximal value.

With the interferer at either $0°$ or $330°$ there is no benefit to be achieved by rotating away from directly facing the target. For the condition with the interferer at $60°$ a small increase in B_{BE+BU}, 0.8 dB, can be achieved with a $20°$ turn towards the interferer, arising from an increase in both the B_{BE} and the B_{BU}. But with the interferer at $90°$ the gain that can be achieved rotating away from the target is much greater, with a $35°$ rotation producing an increase in B_{BE+BU} of 5.1 dB, and even a rotation as small as $10°$ producing an increase in the order of 2.5 dB, whilst a rotation of $20°$ produces an comparable B_{BE+BU} as that of a $30°$ rotation, 12.3 dB compared to 12.7 dB.

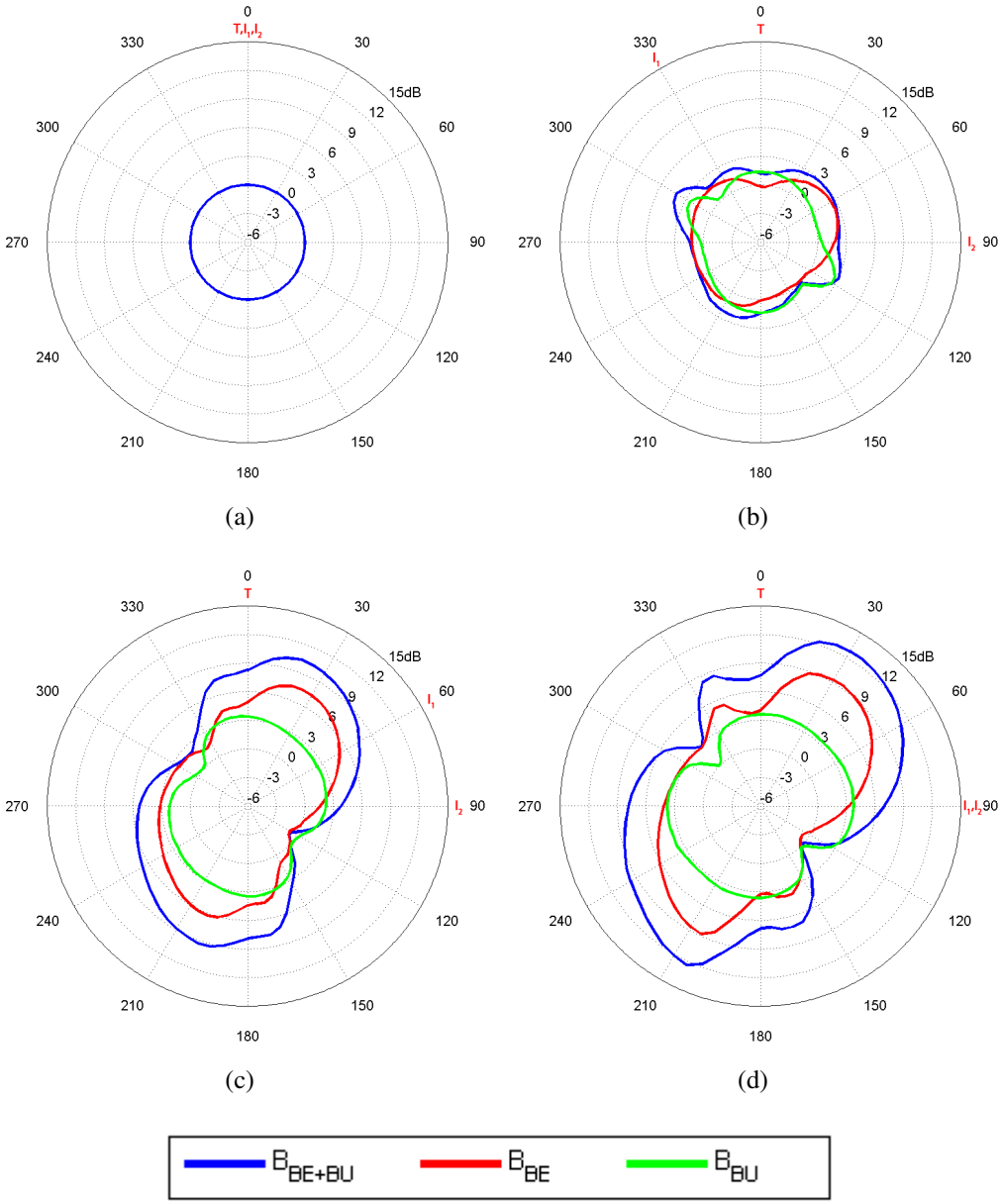


Figure 4.3: Head orientation polar plots, Hawley *et al* (2004) two interferer conditions, with target at 0° and interferers at: (a) 0°, 0°, (b) 330°, 90°, (c) 60°, 90°, (d) 90°, 90°

4.1.2.2 Two interferers

Figure 4.3 shows the results of modelling the four 2-interferer noise conditions used in Hawley *et al*, with the target at 0° and either both interferers at 0° (fig. 4.3(a)), or one interferer at 90° and the other at 330° (fig. 4.3(b)), 60° (fig. 4.3(c)) or 90° (fig. 4.3(d)).

The results of modelling the first condition, with all sources collocated at 0° (fig. 4.3(a)) returns the same constant result as obtained in the comparable single interferer condition (fig. 4.2(a)). Because the variation with azimuth is the important factor, and the absolute values of the modelling results are irrelevant, the B_{BE+BU} values for all conditions were corrected for overall interferer level; 3.01 dB was added to the B_{BE} values, and so to the B_{BE+BU} values. This need for correction arises from the increased power in the interferer due to the additional second interfering source. Consequently, such correction eases comparison between, single, double, and triple interferer conditions.

Looking initially at the B_{BU} for the first of the spatially separated two-interferer conditions, fig. 4.3(b), it can be seen that there is a peak in the benefit at 120° and 300°, caused by both interferers being on the same side of the head, at $\pm 30^\circ$ and $\pm 150^\circ$ respective to the listener. At these locations both interferers have the same ITD, which is different to that of the target, and so the binaural system is able to effectively cancel both of them. The lowest levels of B_{BU}, approaching 0 dB, occur approximately when the listener is facing directly toward or away from one of the interferers, $60^\circ - 240^\circ$ and $330^\circ - 150^\circ$. In these situations the interferer directly in front of the listener can be effectively cancelled, but the second interferer can not be. With respect to the B_{BE} for this condition, the lowest level occurs at 150°. This is where there is one interferer directly behind the listener, which will have no ILD, and the other interferer is at 30° to the left of the listener, which in terms of ILD is approximately equal to that of the target at 150° to the left of the

listener. Thus, there is no benefit available to the listener from using better-ear listening. The maximum B_{BE+BU} available to the listener is 4.1 dB when they are facing $300°$, which is an improvement of 3 dB against facing the target directly.

The results of modelling the second of the two-interferer spatially separated conditions (fig. 4.3(c)) is similar to that produced with a single interferer located at $60°$ (fig. 4.2(c)), except for a reduction in B_{BE+BU} at $0°$ and $180°$. This is due to the addition of the second interferer at $90°$ introducing the effect of the "bright" spot at these points, increasing the level of the interferer at the occluded ear, and so reducing the B_{BE} and therefore the B_{BE+BU}. Whilst the maximum available B_{BE+BU} is similar to that achieved in the single-interferer equivalent, 10.4 dB rather than 10.8 dB, the benefit achieved due to rotation from $0°$ has increased from 0.8 dB to 2.1 dB.

With two interferers present it is possible for a listener to get significant benefits in B_{BE+BU} by rotating away from facing the target directly, up to 3 dB if the interferers are located at $330°$ and $90°$ or up to 2.1 dB if the interferers are at $60°$ and $90°$. For the latter situation this needs only a small rotational movement, of $20°$.

4.1.2.3 Three interferers

Figure 4.4 displays the results of modelling the three-interferer configurations used in Hawley *et al* (2004) with the effect of listener orientation. As with the two-interferer conditions the target is always at $0°$, there are two situations with collocated interferer sources, either all three at $0°$ (fig. 4.4(a)) or all at $90°$ (fig. 4.4(d)), and a further two situations with spatially separated interfering sources, at $330°$, $60°$ and $90°$ (fig. 4.4(b)) or $30°$, $60°$ and $90°$ (fig. 4.4(c)).

In the same way as with the two interferer conditions, the B_{BE} and B_{BE+BU} predicted

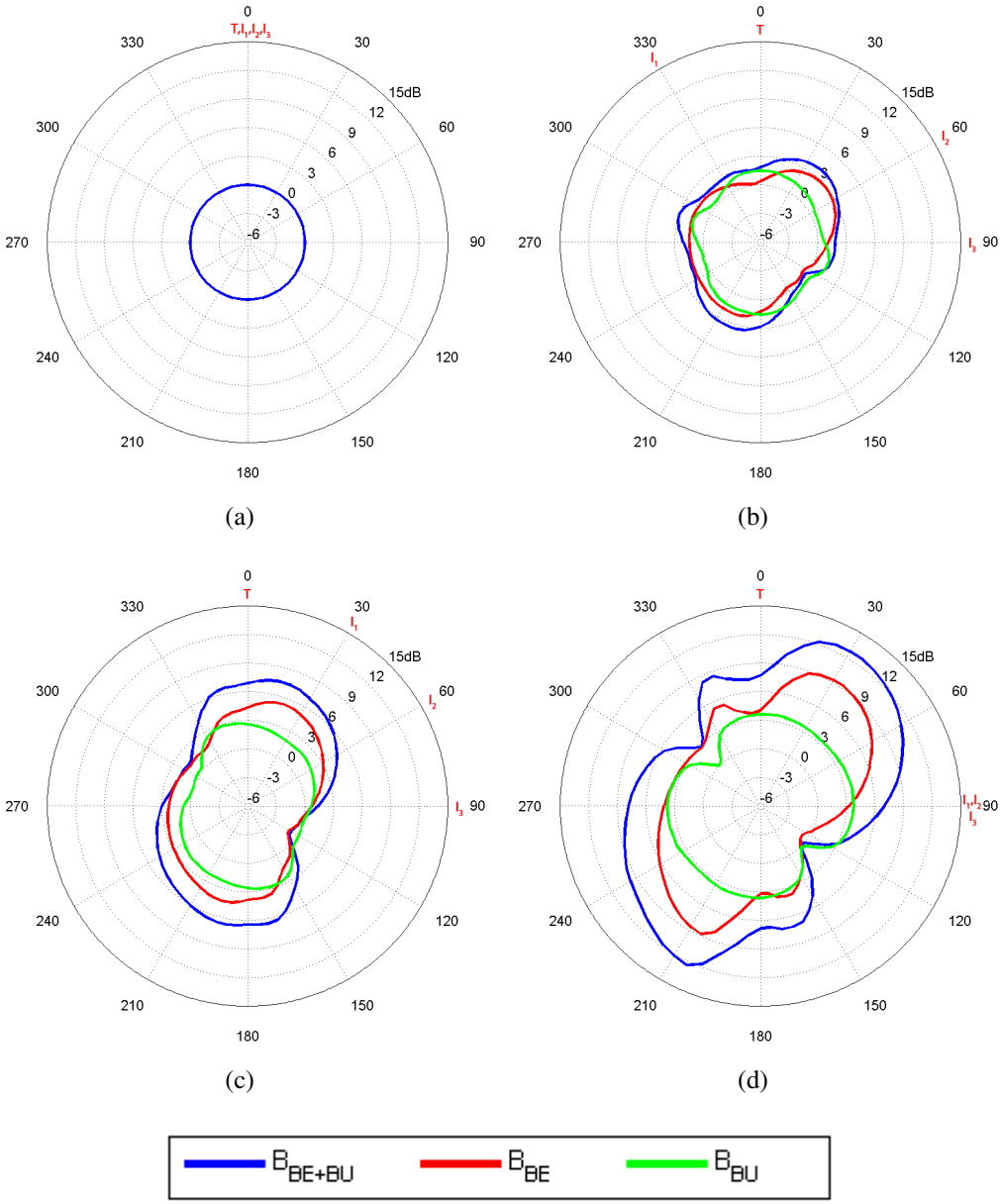


Figure 4.4: Head orientation polar plots, Hawley *et al* (2004) three interferer conditions, with target at 0° and interferers at: (a) 0°, 0°, 0°, (b) 330°, 60°, 90°, (c) 30°, 60°, 90°, (d) 90°, 90°, 90°

values for the three interferer conditions were corrected for the overall interferer level, requiring the addition of 4.77 dB. Consequently, when all interferers are at $0°$ or at $90°$ the results are identical to the single-interferer equivalents (fig. 4.2(a) and 4.2(d) respectively).

The two spatially separated conditions generally follow the same pattern as the equivalent two-interferer conditions. For condition (b), with sources at $330°$, $60°$ and $90°$, there is a reduction in the lobe that was present on the $120° - 300°$ axis. This is because the addition of the third interferer at $60°$ with a different ITD from the other two sources, reduces the B_{BU} that was present with only two interferers. The maximum value obtained is reduced from 4.1 dB to 3.9 dB, occurring at $40°$, but the benefit due to rotation is reduced from 3 dB to 2 dB, due to the initial B_{BE+BU} when the listener is facing the target increasing by 0.7 dB.

The second spatially separated condition (fig. 4.4(c)) follows the same pattern as the equivalent two-interferer condition, but as with the previous condition it is reduced in maximum level. With two interferers the listener could achieve a B_{BE+BU} of 10.4 dB, whilst the addition of a third interferer at $30°$ reduces this to 7.5 dB, though it occurs at the same azimuth of $20°$. The benefit due to rotation of the listener is reduced from 2.1 dB with two interferers to 0.6 dB with the addition of a third interferer.

With the addition of a third interfering source, the benefit to the listener of rotating their head is reduced when the sources are spatially separated, to 2 dB for the configuration tested with interfering sources positioned either side of the target, and to 0.6 dB when all interferers are to the right of the target. In both instances this loss of benefit is caused by a reduction in the ability of the listener to use better-ear listening and binaural unmasking.

4.2 Optimum-Orientation Spatial Maps

In chapter 3 a number of room maps were generated, looking at the effect of listener position, the level of reverberation, and the number of interferers on the ability of a listener to understand target speech in a background of noise. In those maps the listener was always facing the target, but as can be seen from the previous polar plots, this may not always be the optimum head orientation, with predicted $\text{TIR}_{\text{BE+BU}}$ levels increasing by up to 5 dB for certain configurations. By including the ability of the listener to rotate their head, at each point in the room, it is possible to look at the effect of orientation over a greater number of positions, with varying source distances and relative azimuths.

To do this, four sets of maps were generated, again based upon the configurations of Hawley *et al* (2004). Figure 4.5 shows the plan of the 10 m long by 6.4 m wide room that was used, which has a ceiling height of 2.5 m. Within the room a target was placed at 5 m (l) by 5 m (w), and three interferers placed at 4 m (l) by 4.74 m (w), 6.74 m (l) by 4 m (w) and 7 m (l) by 3 m (w). These positions were chosen so that a listener at 5 m by 3 m would experience the same source azimuths as those used in the Hawley *et al* (2004) experiments, and be comparable to the polar plot shown in figure 4.4(b). The four sets of maps generated are three sets using only one of the interferers each, and one set using all three interferers. For each set, three maps are generated, the first being comparable with the better-ear and binaural unmasking maps generated in chapter 3, with the listener facing the target, and having the use of both better-ear listening and binaural unmasking. The second map is the same as the first, but this time with the listener facing the optimum direction. This is modelled by rotating the listener through 360° in 5° increments at each listening position, and the $\text{TIR}_{\text{BE+BU}}$ calculated for each direction. the optimum orientation is the one with the maximum of the 72 values predicted. The

final map shows the benefit of listener rotation as a function of position, calculated by subtracting the facing-the-target map from the optimum-orientation map. All of these maps are generated in an anechoic space, so there are no effects of the room.

4.2.1 Single Interferer Configurations

4.2.1.1 Interferer 1

The first set of maps used the "Int 1" interferer position from figure 4.5. Figure 4.6, shows (a) the TIR_{BE+BU} when the listener is facing the target, (b) the TIR_{BE+BU} when the listener adopts the optimum head orientation and (c) the benefit of optimising head orientation, compared to facing the target. The map of the listener facing the target follows the same pattern as the single interferer anechoic maps generated previously (See Ch. 3.1.1, p. 55). For the optimum direction map the colour scale indicates the level of the predicted TIR_{BE+BU}, whilst the arrows denote the optimum direction for the listener to adopt in order to attain that level. The position that the arrow relates to is at the origin of the arrow, spaced on a 30 cm by 30 cm grid.

For the optimum-orientation map there are three distinct regions of optimum orientation, 1) the area on the target-interferer axis, 2) the areas near to the target, and 3) the distant areas. The most complex area, with the least regular pattern is that of the target-interferer axis. On this axis there is often a large change in direction between adjacent points. For points close to the target and not on the target-interferer axis the optimum direction is to look either directly at the target, or between the target and interferer. These orientations maximise better-ear listening, either by distributing the target and interferer to opposite sides of the head, or as with the 5 m by 3 m position, to look directly at the target because, as shown in the polar plot (fig. 4.2(b)), there is no benefit from rotating

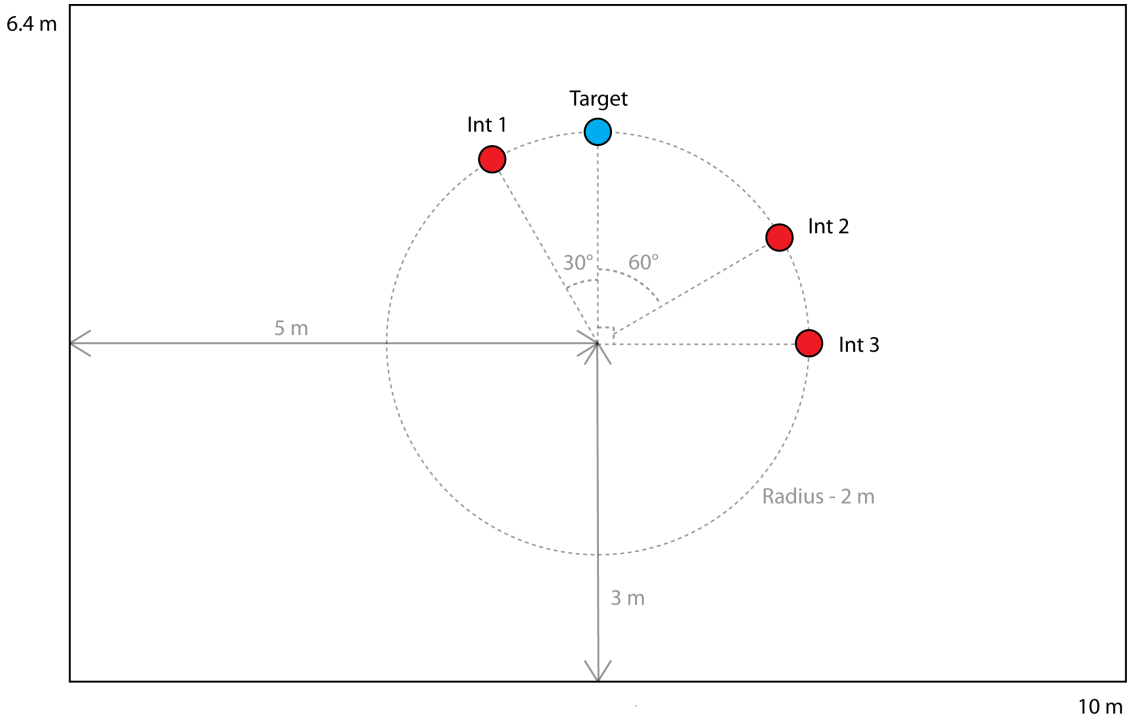
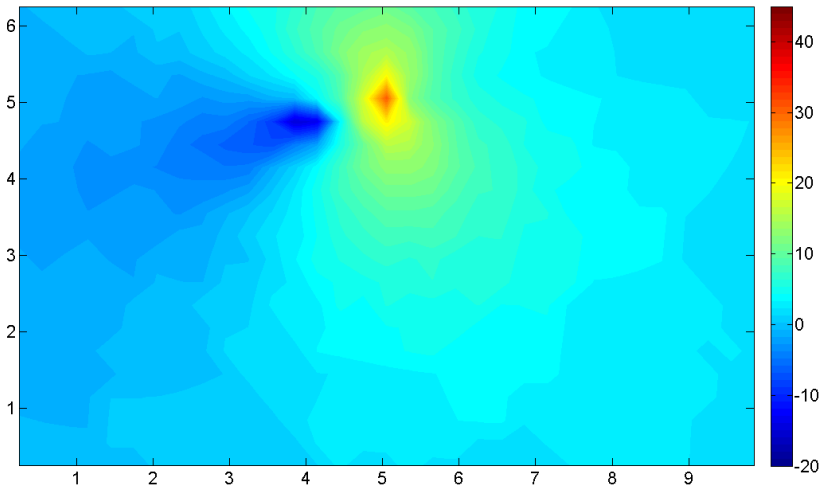
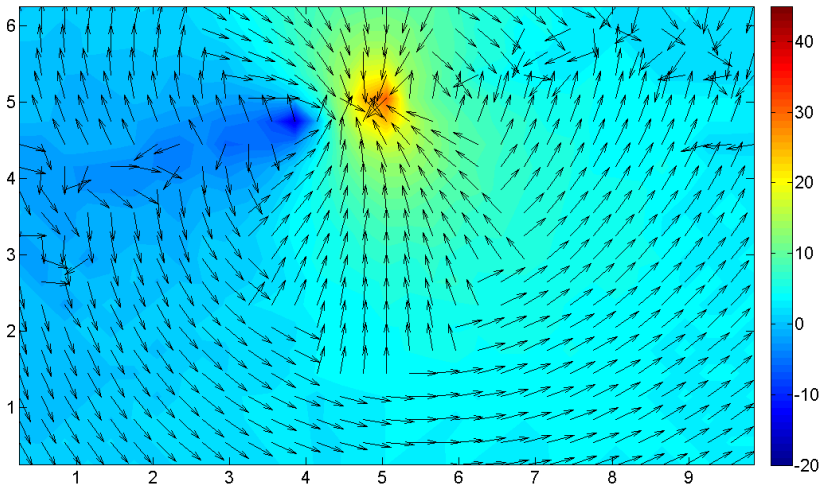


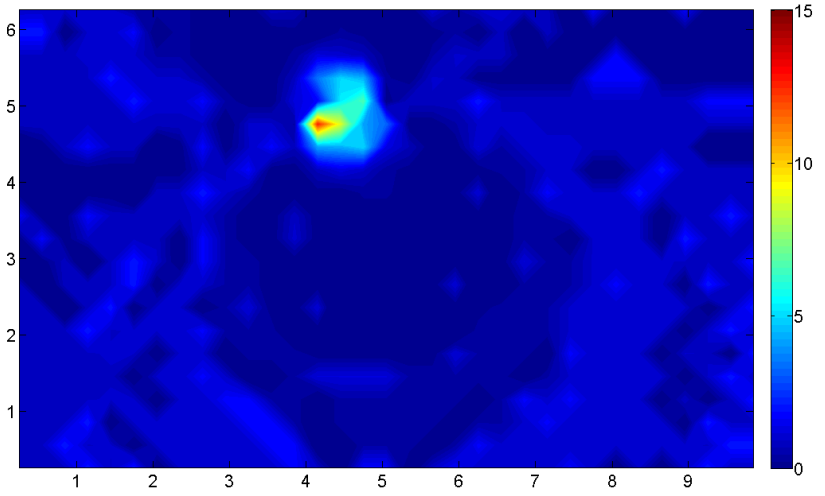
Figure 4.5: Plan of the room and sources used in the generation of the room maps with the effect of head rotation. Source positions are chosen so as when listener is stood at 5 m x 3 m they match the configurations used by Hawley *et al* (2004). As the room is anechoic there is no concern for standing waves that may have been present at the central listening space if the room was reverberant, and also the room dimensions / walls have no effect on the prediction, but are merely included as a bounding box for the analysis grid.



(a)



(b)



(c)

Figure 4.6: Room maps, using Int 1, showing TIR_{BE+BU} when (a) listener is facing the target, (b) listener faces the optimum direction, and (c) the benefit of head rotation.

the head. At more distant locations the optimum head orientation is at approximately $90°$ to the target. To gain insight into the reason that this is the optimum orientation, Figure 4.7 shows the polar response of one of these positions. The listener in this instance is at 6.55 m by 1.45 m, and the results are shown as (a) the B_{BE+BU}, B_{BE}, and B_{BU}, and (c) the individual left and right ear TIRs. For these polar plots, the $0°$ orientation is when the listener is facing across the width of the room (up the 6.4 m axis of the room plots), whilst the target is at $336.5°$, and the interferer is at $322.2°$. At this position the optimum orientation is at approximately $68°$ from the vertical axis, at a level of 4.4 dB. From panel (a) it can be seen that the peak of TIR_{BE+BU} at the optimum orientation is due to the TIR_{BE} with little contribution from binaural unmasking. Breaking the TIR_{BE} down into the left and right ears, panel (b), shows that it is the right ear that has the greater TIR. When the listener is facing $68°$ the target is at $90°$ to their left hand side. This means that the target is effectively in the "bright-spot" of the listeners right ear, whilst the interferer is occluded from the right ear by the head. This gives a greater TIR than at the left ear where both sources are unoccluded, even though the relative levels at the right ear are reduced.

As discussed in chapter 2 (see 2.3, page 42), the HRTFs of the KEMAR manikin, as produced by Gardner & Martin (1994), are highly symmetrical which may have unduly enhanced the effect of the "bright-spot". As such the modelling for this single position was replicated using human-HRTFs of subject "mer", taken from the AUDIS catalogue (Blauert et al. 1998), the result of which is shown in figures 4.7(b) and 4.7(d). Broadly speaking the modelling using the human HRTFS matches the results of the modelling using the KEMAR HRTFs, albeit with slightly more "noise". The key difference being that whilst the peak in B_{BE+BU} predicted at $68°$ is still present, there is also a peak present at $350°$ which is not present in the KEMAR-based modelling. This peak is slightly greater

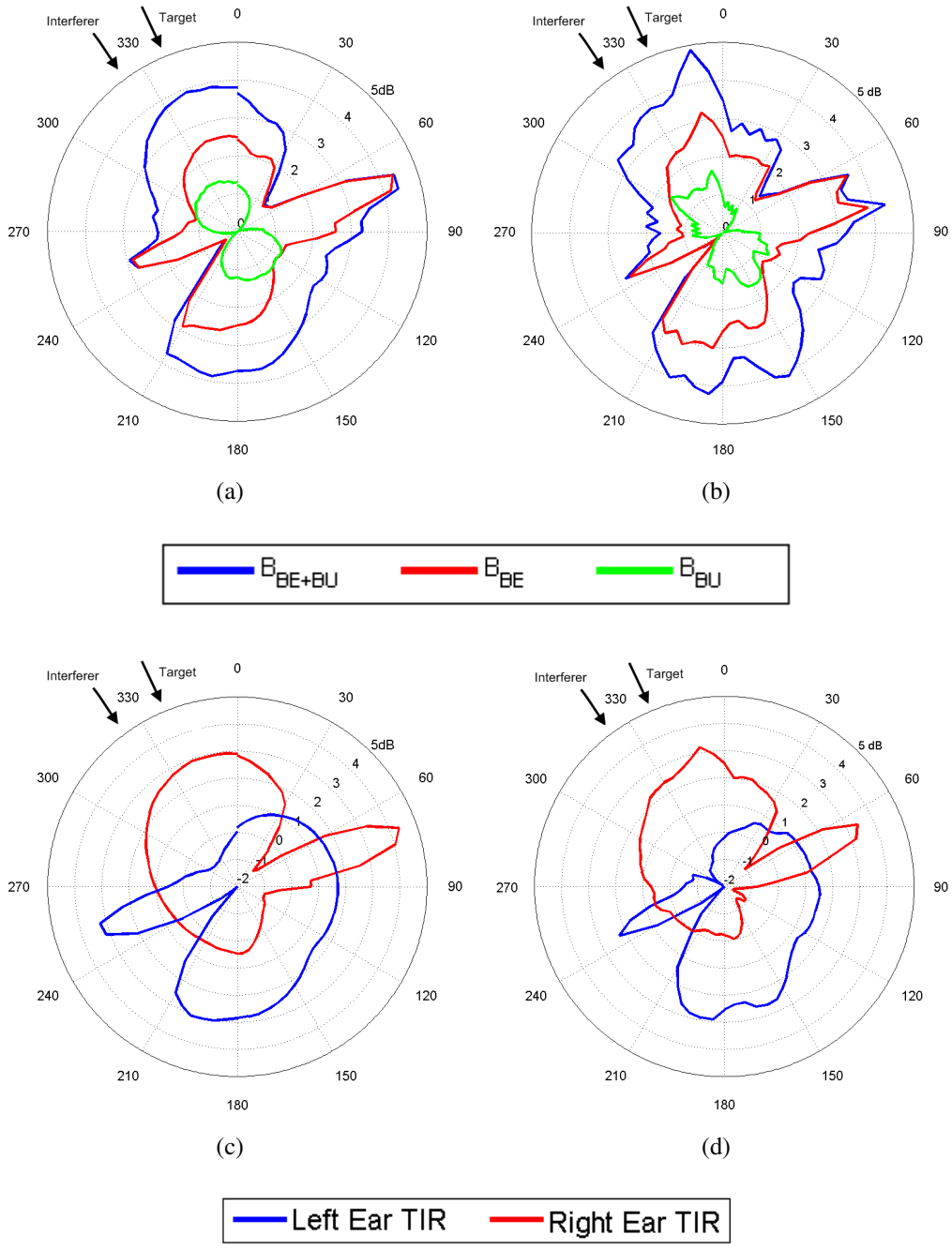


Figure 4.7: Head orientation polar plots, detail of position 6.55 m x 1.45 m for "Int 1" Hawley *et al* (2004) head optimisation map. Target is at 336°, Interferer at 322°. Showing benefit of better-ear listening and binaural-unmasking for either (a) a listener with KEMAR manikin HRTFS, or (b) with human HRTFS (subject "mer" from the AUDIS catalogue (Blauert et al. 1998)) Also showing left and right ear TIRs for (c) the KEMAR-based listner, or (d) the human listener.

than the peak at $68°$. This peak is caused by both differences in the predicted left and right ear TIRs than those predicted for the KEMAR manikin, but also a greater predicted $B_{BU}$when using the human HRTFs than using the KEMAR HRTFs. These differences are caused by the slight variations between the human and the KEMAR manikins, and will vary from subject to subject, but also may be a factor of the low spatial resolution of the human HRTFs, which are only sampled every $15°$ in azimuth, and the interpolation of all points between the measured azimuths.

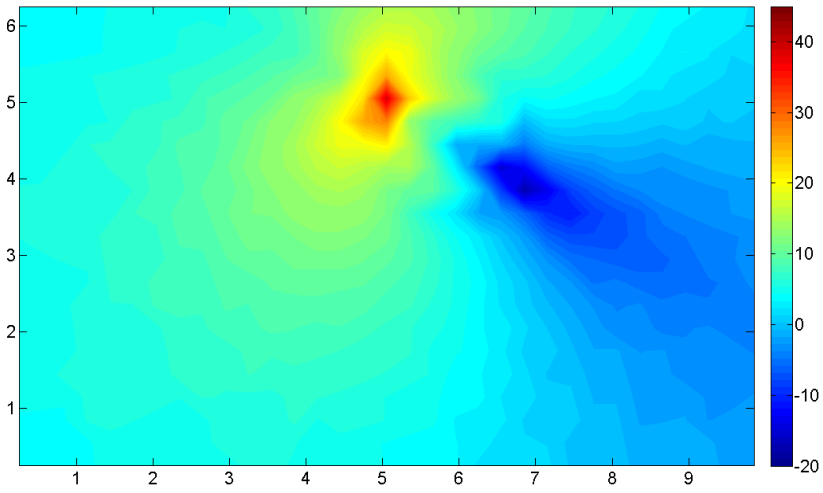
Figure 4.6(c) shows the benefit due to head orientation. The maximum benefit available is 12.6 dB, occurring in the space between the target and interferer. Out of the 693 positions in the analysis grid 535 of them have benefits due to head orientation, 100 of them greater than 1 dB, giving an overall mean of 0.75 dB across all positions.

4.2.1.2 Interferer 2

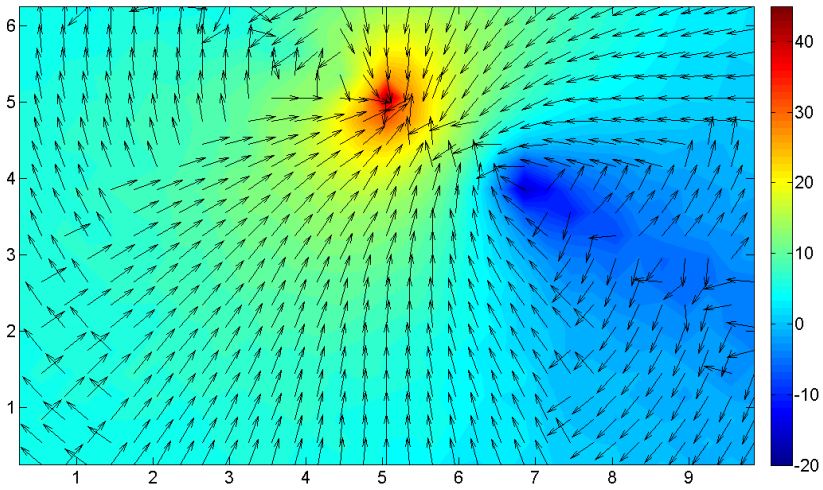
Figure 4.8 shows the results of modelling the effects of head orientation with a single interferer located at the "Int 2" position. Again, the pattern produced when modelling the listener as facing the target follows the same principles as the single-interferer anechoic maps produced in chapter 3.

With respect to the optimal head orientation, there are still the three distinctive regions shown in the previous maps, the on-axis area, the area where it is best to approximately face the target, and the area where it is best to face $90°$ to the target. The difference in this case is that the number of locations where it is most beneficial to face either directly towards, or slightly between the target and the interferer, is increased, and the areas in which it is best to face at $90°$ to the target is decreased.

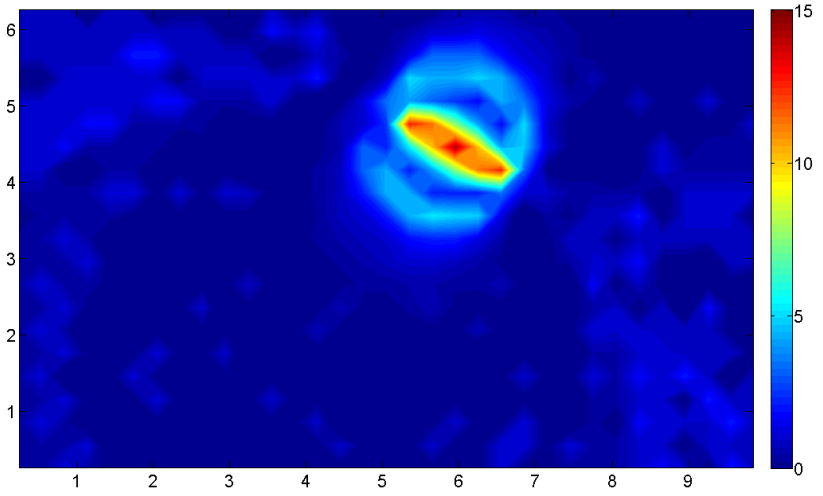
Because the target and the interferer are spaced further apart than when using the "Int



(a)



(b)



(c)

Figure 4.8: Room maps, using Int 2, showing TIR_{BE+BU} when (a) listener is facing the target, (b) listener faces the optimum direction, and (c) the benefit of head rotation.

1" interferer position, there is a clearer pattern of on-axis benefit produced (fig. 4.8(c)). Approximately 420 of the 693 listener positions tested produce benefits due to head rotation, with a maximum benefit of 14.3 dB. Yet as the high levels of benefit are confined to the area close to the target and the interferer, there are less than 40 locations where benefits greater than 3 dB occur, and the mean value across the room is still 0.75 dB. The high peaks of benefit occur when the listener is on the axis between target and interferer, where if they face the target directly there will be no ILD or ITD produced, meaning the TIR_{BE+BU} will be influenced by the effect of source distance only. In this situation head rotations will be highly beneficial in increasing listener performance. The second major area of benefit of head rotation is the circle that passes through the target and the interferer. In this area, when the listener is facing the target the interferer will be at 90° to them, and so benefits of head rotation occur when the interferer direction is shifted away from this value, as shown in 4.2(d).

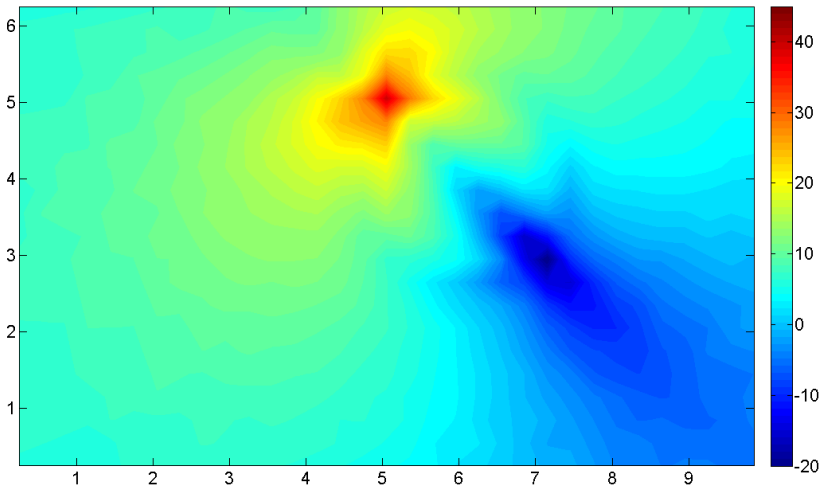
4.2.1.3 Interferer 3

The modelling results of "Int 3" are displayed in figure 4.9. The general pattern of TIR_{BE+BU} for the situation where the listener is facing the target (panel (a)) follows the same pattern as that produced when using "Int 2", albeit rotated and enlarged.

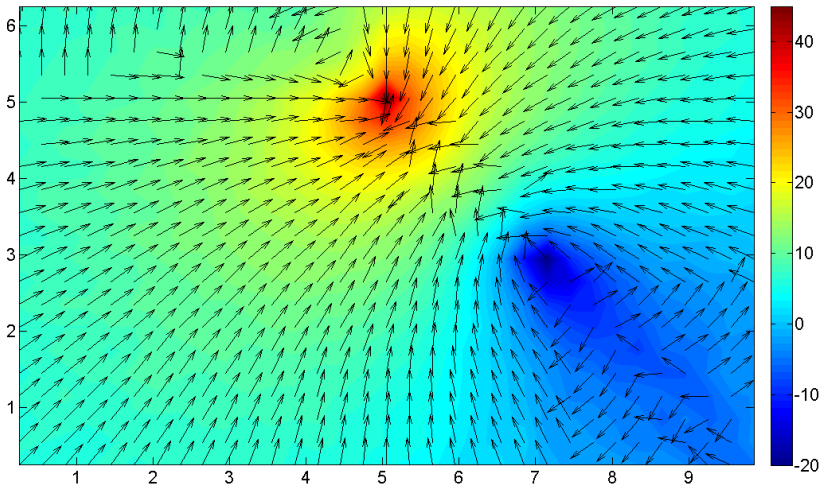
The area in which it is beneficial to face approximately in the direction of the target is increased in size, including almost all of the room other than the on-axis points. The number of listener locations that produce a benefit due to head rotation are increased over using "Int 2", from 420 to 462, with 73 of them producing benefits greater than 3 dB, and with the mean benefit increasing to 1 dB. As with using "Int 2" the main area of benefit is when the listener is on-axis with the target and interferer, peaking at 13.7 dB, as well as

when the listener is on the circle linking target and interferer.

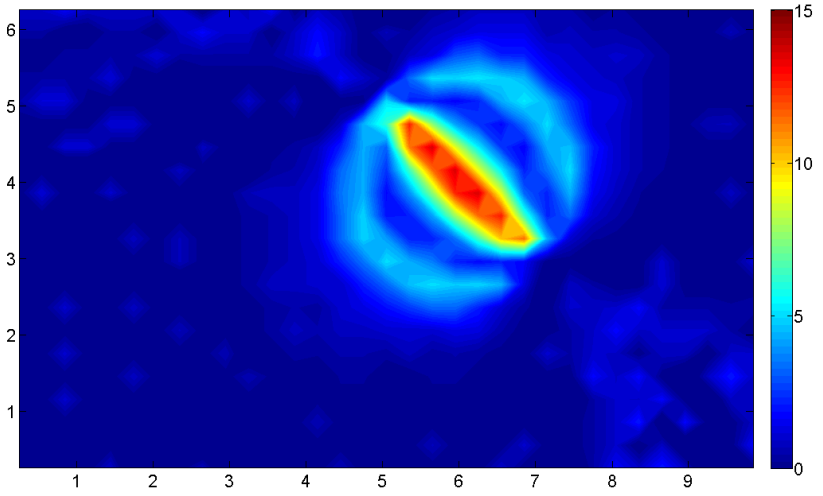
The main effect of interferer position is thus to vary the area in which large benefits due to head rotation are obtainable, increasing when the distance between target and interferer increases.



(a)



(b)



(c)

Figure 4.9: Room maps, using Int 3, showing TIR_{BE+BU} when (a) listener is facing the target, (b) listener faces the optimum direction, and (c) the benefit of head rotation.

4.2.2 Multiple Interferers

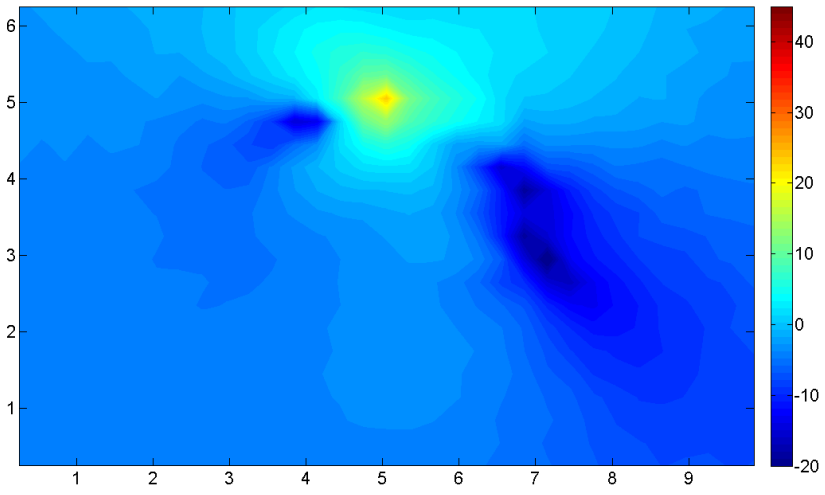
To look at the effect of head orientation when there are multiple interfering sources present all of the interferer positions outlined in figure 4.5 were modelled simultaneously. The effect of head orientation was modelled in the same way as that used in the single interfering conditions, by producing BRIRs for each source at each listener location with the interferer rotating in 5° increments.

With three interferers the $\mathrm{TIR}_{\mathrm{BE+BU}}$ when the listener is facing the target produces the pattern shown in figure 4.10(a). This is similar to the the patterns produced in the modelling of multiple anechoic interferers examined in chapter 3.2.1, when including the effects of both better-ear listening and binaural unmasking, which is effectively the same situation, but with differing source positions. The optimum direction for the listener to face follows a more complex pattern than those observed in the single interferer conditions, as shown in figure 4.10(b). The effect of target-interferer axis is still present, though reduced in terms of the area that each individual axis affects, limiting it to the areas that are close to one of the individual interfering sources.

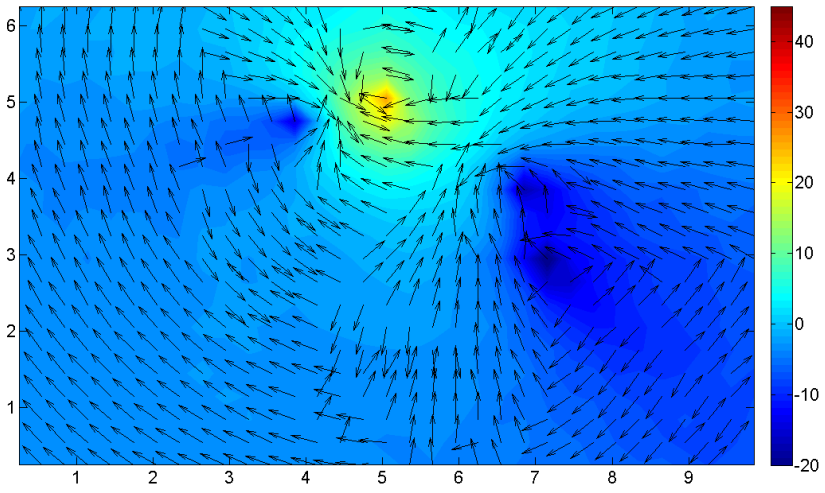
The maximum observed benefit due to head rotation is 11.9 dB, but this only occurs in very close proximity to one of the interferers. The mean benefit is 1 dB, with 615 out of the 693 listener locations modelled producing a benefit, 37 of which are greater than 3 dB.

4.3 Discussion

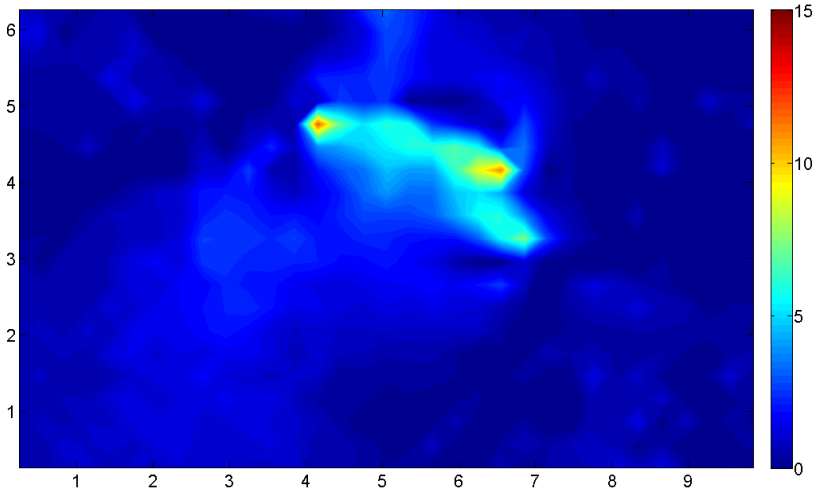
In this chapter two main themes of research were investigated; the effect of azimuthal separation of target and interfering sound sources on the ability of listeners to understand



(a)



(b)



(c)

Figure 4.10: Room maps, using all three interferers, showing TIR_{BE+BU} when (a) listener is facing the target, (b) listener faces the optimum direction, and (c) the benefit of head rotation.

speech in noise, and the benefit to be gained by listeners from rotating their heads when listening to a target in the presence of noise. This was done through modelling listener polar responses to certain configurations, and the benefit obtained by using head rotations.

4.3.1 Polar Plots

Section 4.1.1 looked at the effect of azimuthally separating the target and the interfering source, which is to give a benefit to the listener, approaching 14 dB, depending on the way in which the sources are separated (fig. 4.1). To attain this level of benefit the listener must be able to move the target and the interfering sources to optimal azimuths individually, which is unlikely.

The more realistic situation is that modelled using the Hawley *et al* (2004) configurations, Section 4.1.2, where the listener is able to rotate their head, but all sources remain stationary, so that they all undergo the same change in azimuth with respect to the listener. In this instance benefits can still be achieved, up to a maximum of 5.1 dB in the modelled conditions. One issue with these benefits is that sometimes they can require the listener to make large head movements, such that they have to face away from the target. Aside from breaking the social convention that a listener faces the person that they are attending to, the listener may also lose the ability to use visual cues. These are cues that aid understanding of speech, particularly lip movements (Reisberg, McLean & Goldfield 1987), tongue movements (Badin, Tarabalka, Elisei & Bailly 2010) and head or eyebrow movements (Foxton, Riviere & Barone 2010). The loss of such visual cues may outweigh the benefit obtained by the auditory system in real-world situations. If the amount of rotation is limited to a maximum of $20°$, then the maximum predicted benefit is 2.5 dB in the conditions modelled here, occurring when the target is at $0°$, the interferer at $90°$, and the

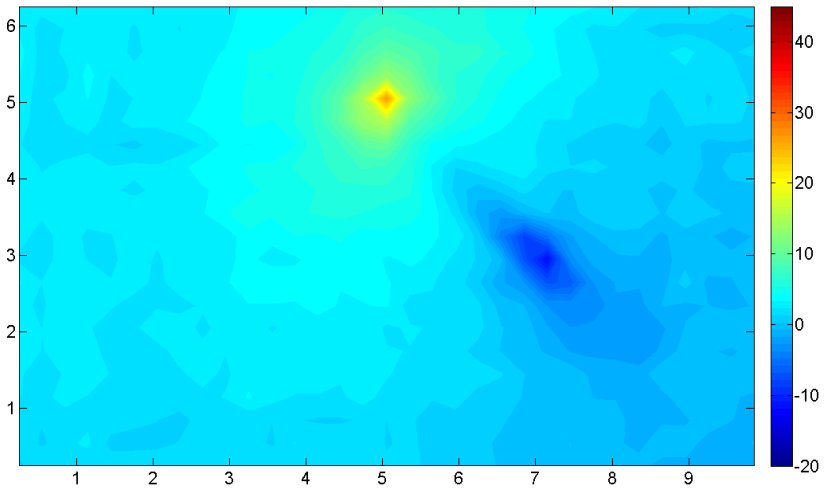
listener faces 20°, the Hawley *et al* (2004) polar configuration shown in figure 4.2(d).

4.3.2 Spatial Mapping

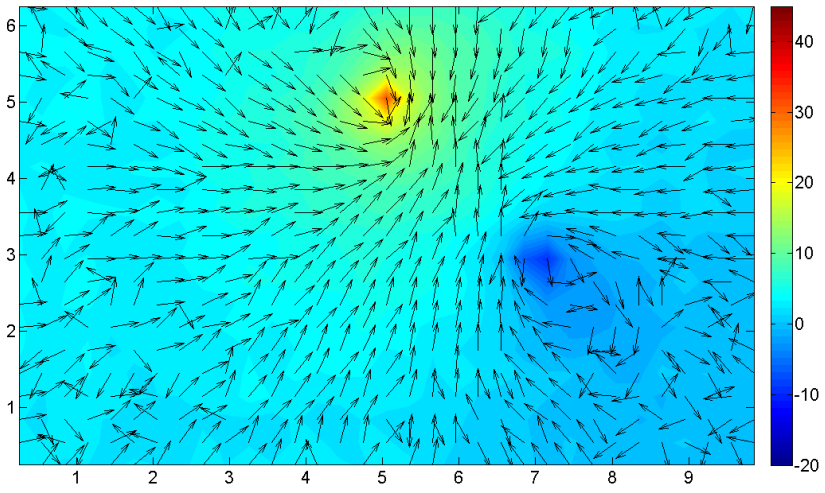
The spatial mapping showed that head rotations can be beneficial to the listener, depending on their location within the room. The areas that benefit the greatest from head rotations are those that are close to both the target and the interferer, and in particular on the target-interferer axis. The majority of locations within each room configuration tested produced benefits through head rotations, with a greater spacing between the target and interferer increasing the size of the area in which the high levels of benefit occur.

4.3.3 Limitations

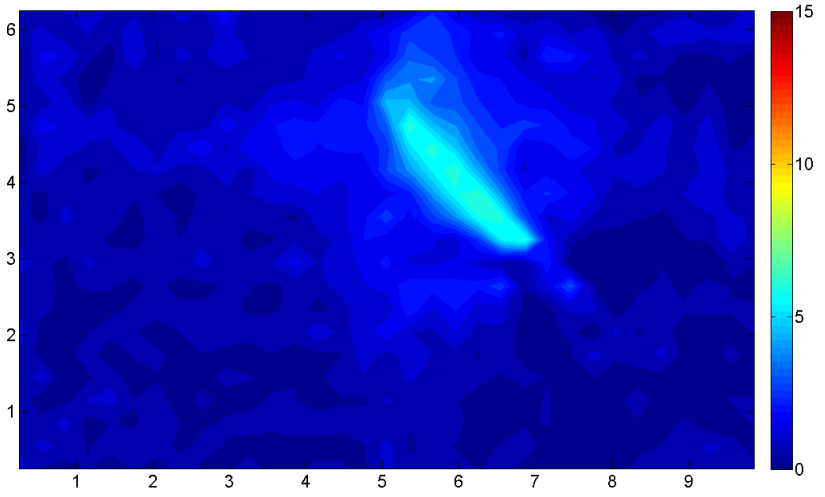
One of the limitations of the modelling in this chapter is that in all instances the space is anechoic. The effect of this is to portray the "best case" scenario, in that it will give the highest possible benefits due to head orientation. The addition of reverberation will decrease the ILDs and ITDs. It will also decrease the B_{BU}, as previously shown in chapters 2 and 3, in part due to the reduction of the ITDs, but also due to the reverberation decreasing the interaural coherence of the interferer. Given that the areas in which optimal head orientations benefit the listener are those areas that are close to the target and interferers, which in turn are the areas with high direct-to-reverberant ratios, there will still be a benefit available to the listener when using head orientations. An example of this is shown in figure 4.11, which models the "Int 3" position, as used in section 4.2.1.3, with all surfaces in the room having an absorption coefficient of 0.3. The addition of reverberation in the initial condition where the listener faces the target (fig. 4.11(a)) is to reduce the range of observed values, decreasing the standard deviation from 7.2 dB in the



(a)



(b)



(c)

Figure 4.11: Room maps, using Int 3 including the effects of reverberation, showing TIR_{BE+BU} when (a) listener is facing the target, (b) listener faces the optimum direction, and (c) the benefit of head rotation.

anechoic condition (fig. 4.9(a)) to 2.7 dB. More importantly, the maximum observed benefits due to head orientation are reduced, 6.6 dB compared to 13.7 dB, but the mean value remains the same, indicating that whilst those areas that get large benefits are reduced, the number of areas where benefits are available is increased. 689 of the 693 positions tested have some benefit available thanks to optimal head orientation, with 79 of the positions having benefits in excess of 2 dB. Whilst the anechoic cases may be the best case scenario for showing maximum benefits, in the presence of reverberation optimal head orientation affects larger areas of the room.

A second limitation of the modelling done so far in this thesis is that it has always assumed to be for listeners with normal hearing, whilst those with a hearing impairment are in fact the ones that are going to struggle the most with understanding speech in noise. One issue with hearing loss, particularly for persons using Cochlear Implants (CIs), is the loss of the ability to use binaural unmasking due to the way CIs encode the audio signal. As such the only benefits available to them will be those arising from better-ear listening.

4.3.4 Conclusion

By optimising the orientation of their head, listeners can make substantial benefits through maximising the ILDs and ITDs of the target and the interfering sources, be it with one, two, or three interferers, in anechoic or in reverberant spaces. When using a method to predict the intelligibility of speech in the presence of a noise interferer such benefits must be included, if an accurate portrayal of the environment is to be achieved.

Chapter 5

Audiological Implications

In Chapter 4, the effects of head orientation on the intelligibility of speech in noise was examined. As part of this examination polar plots of the effects of independently moving one or other source about the listener were generated (fig. 4.1, p. 84), but only with reference to the overall B_{BE+BU}, in order to model normally hearing listeners. For people with hearing impairment different patterns of benefit may be produced, depending on the specific nature of their individual impairment. For some it may be just an overall reduction of the predicted levels, whilst for others, specifically Cochlear Implant users, there will be a total loss of the benefit due to binaural unmasking, and if, as is common, they have a unilateral implantation, a reduction of the benefit of better-ear listening as well.

5.1 Effect of Orientation on CI Users

5.1.1 Principles and Limitations of CIs

Cochlear Implants (CIs) are prosthetic devices used to restore hearing in severely or profoundly deaf individuals. This is achieved by surgically implanting an electrode array so that it electrically stimulates the cells of the auditory nerve fibres from inside the scala tympani, one of the sections within the cochlear. This is linked to a processing unit behind the person's ear, which encodes the acoustic information received by a microphone into electric pulses to be fed to the array. Dependent on the particular electrode array implanted there are between 12 and 24 electrodes, used to stimulate different auditory nerve fibres, corresponding to different frequency bands. For the understanding of speech in quiet as few as four electrodes are required to facilitate a high level of comprehension (Shannon, Zeng, Kamath, Wygonski & Ekelid 1995). But, Qin & Oxenham (2003) showed that in the presence of noise the level of performance drops rapidly with reduced numbers of encoding channels, at least for simulated implantation tested on normal hearing listeners. Dunn, Tylet, Witt & Gantz (2006) also showed significant differences in ability due to the number of channels used for the encoding of speech with a babble noise interferer.

One of the limitations of cochlear implants is the way they encode the audio information. An acoustic waveform can be defined as two sections, the temporal fine structure, and the envelope (fig. 5.1). The temporal fine structure is the cycle-by-cycle variation in air pressure, while the envelope is the change in amplitude of this waveform over time. Some CIs only encode the envelope of a sound (Wilson, Finley, Lawson, Wolford, Eddington & Rabinowitz 1991), though research is investigating the possibility of also encoding the temporal fine structure (Majdak, Laback & Baumgartner 2006, van

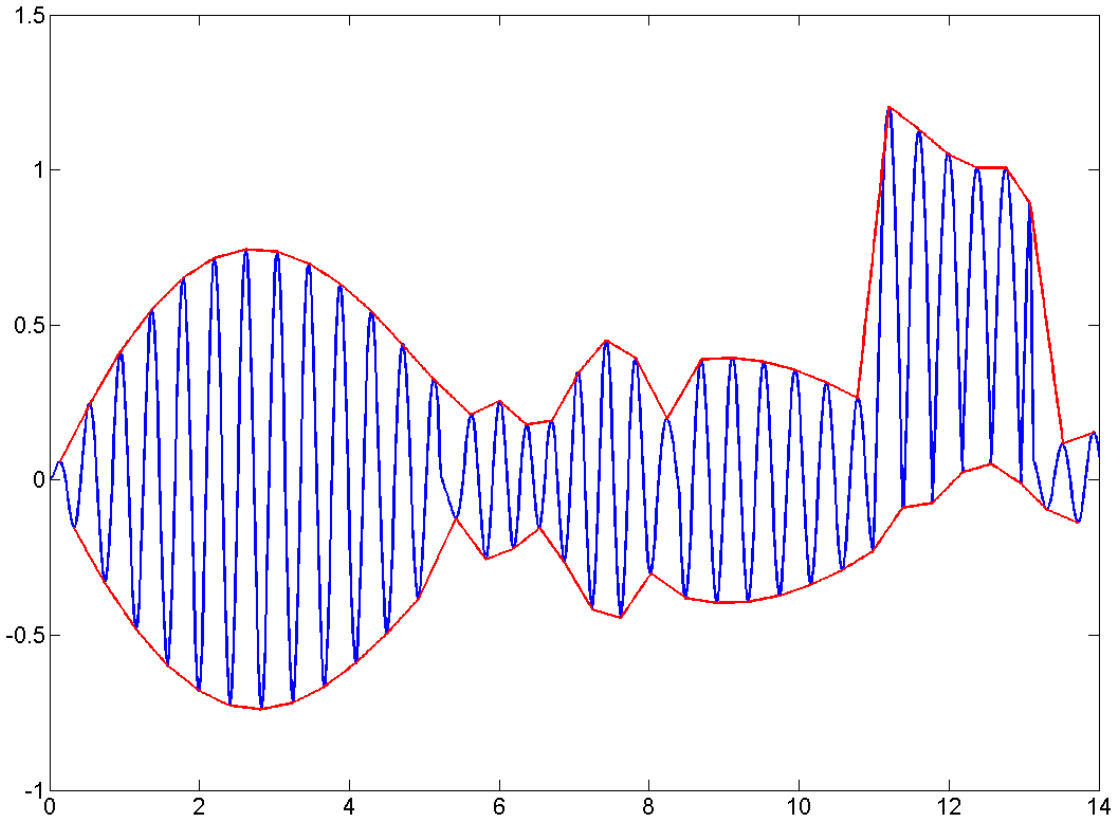


Figure 5.1: Example of temporal fine structure of an acoustic signal, in blue, and the corresponding envelope, in red. Time across the x-axis, amplitude on the y-axis.

Hoesel 2007) and this is being applied in more recent CIs. Whilst binaural unmasking is not reliant on the fine structure of the signal, and has been found for envelope cues in normally hearing listeners (van de Par & Kohlrausch 1997), it has been shown to be of limited use for cochlear implant users (Long, Carlyon, litovsky & Downs 2006).

The National Institute for Health and Clinical Excellence (NICE) recommend that patients with severe to profound deafness receive unilateral (single) implantation, rather than bilateral implantation, which they only recommend for children (who are albeit the largest group of CI users) or persons who are blind as well as deaf, increasing their reliance on the auditory information (NICE 2009). This means that for many CI users who have been implanted through the NHS even the benefit of better-ear listening is removed, limiting the user to simple head shadow benefits when the target and interferer are in beneficial locations.

5.1.2 Benefits of Head Orientation for CI Users

A number of studies have examined the benefits of bilateral cochlear implantation in understanding speech in noise. These studies appear to confirm a relatively small benefit of binaural hearing compared to normal hearing listeners (van Hoesel & Tyler 2003, Loizou, Hu, Litovsky, Yu, Peters, Lake & Roland 2009, Schleich, Nopp & D'Haese 2004, Muller, Schon & Helms 2002, Tyler, Gantz, Rubinstein, Wilson, Parkinson, Wolaver, Preece, Witt & Lowder 2002, Buss, Pillsbury, Buchman, Pillsbury & Clark 2008, Lovett, Kitterick, Hewitt & Summerfield 2010, van Deun, van Wieringen & Wouters 2010). But all of these studies measured the ability of participants when the target was located directly in front of them, $0°$, and the interferer was either collocated at $0°$, or to one side at $90°$. As has been shown in chapter 4, this configuration is not the most advantageous of

positions for normally hearing people.

Figure 4.1 (p. 84) showed the predicted B_{BE+BU} as a function of the source azimuth of the target and the interferer. By breaking the configurations down into the constituent parts, the TIR_{BE} and B_{BU}, we can see the predicted responses for bilateral cochlear implantees, who will have access only to the benefit of better-ear listening, as shown in figure 5.2.

In all of the empirical studies, the benefit of spatially separated sources was measured with the target at $0°$ and the interferer at $90°$, this is equivalent to the $90°$ position on panel 5.2(a). At this position, the predicted TIR_{BE} for a bilateral cochlear implantee will be 4 dB. If the studies had measured the participants with the interferer located at $70°$, with the target still in front, a larger benefit would have been predicted of 6.7 dB. Greater benefits can be shown by having neither source located in front of the listener, with predicted values of up to 10.4 dB when the sources are located $\pm 65°$ (fig. 5.2(c)). This suggests that while previous studies of the benefits of bilateral cochlear implants showed moderate improvements when the sources were spatially separated, much greater improvements could be found by changing the positions of the sources that were modelled. Also, as with the modelled effects of head orientation by normal hearing people carried out in section 4.1.2, bilateral CI users may be able to maximise better-ear listening effects by using simple head movements, which in the previous studies was either discouraged (van Hoesel & Tyler 2003, Tyler et al. 2002), prevented by use of a head rest (Schleich et al. 2004), or made impossible by passing virtual-acoustic stimulations directly to the electrode array (Loizou et al. 2009, Buss et al. 2008). It may be that experienced bilateral CI users are making use of better-ear listening through head movements in natural situations, which might go some way to explain the high level of satisfaction they report (Noble, Tyler,

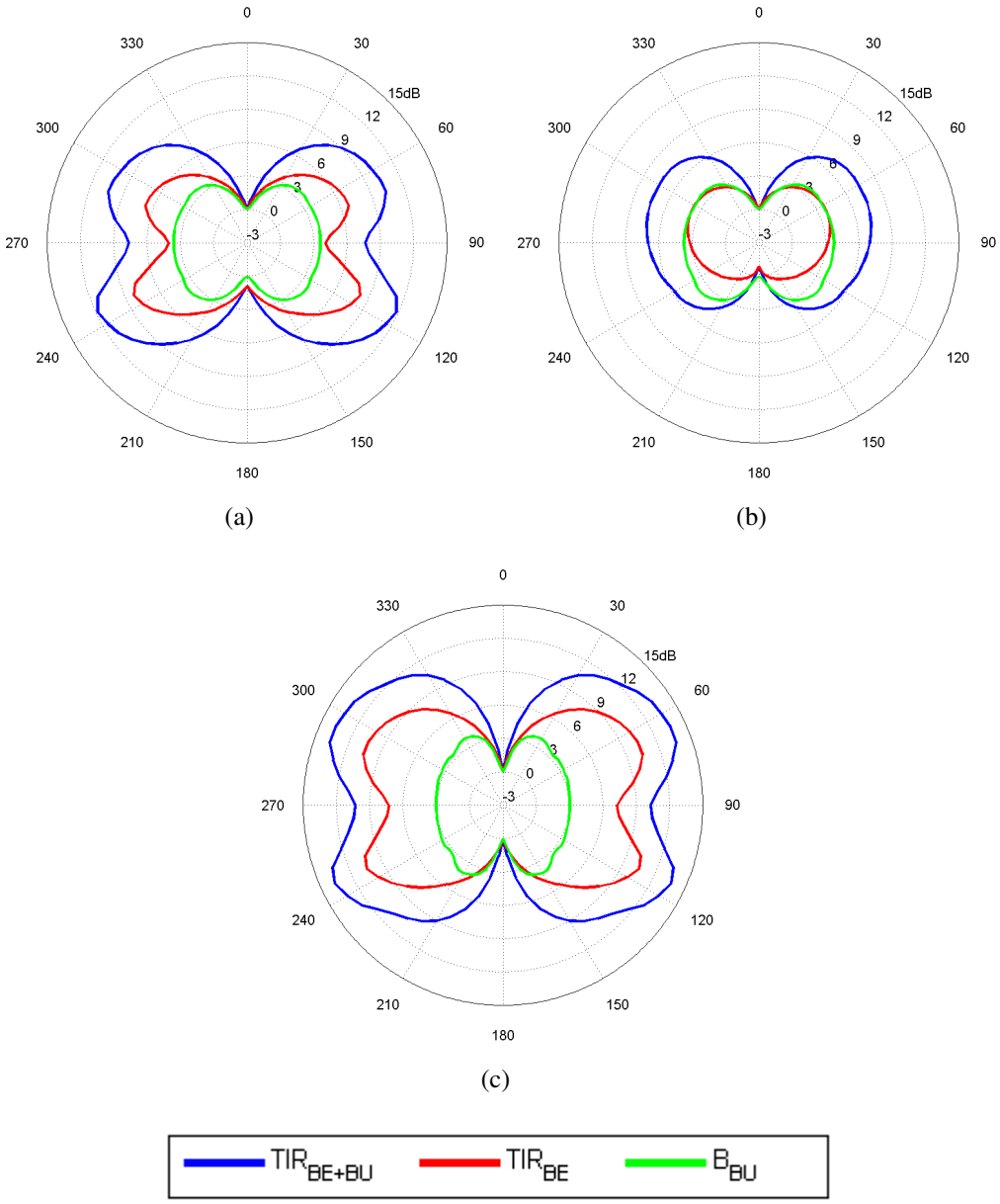


Figure 5.2: Polar plots of the modelled TIR_{BE+BU}, TIR_{BE} and B_{BU} with respect to source azimuth, either having (a) a target fixed at $0°$ and a moving interferer, (b) an interferer fixed at $0°$ and a moving target, or (c) with neither source fixed, both moving around the listener in equal and opposite directions.

Dunn & Bhullar 2008), which is unexplained by the empirical studies.

For any given spatial configuration, bilateral cochlear implantees will have the ability to use better-ear listening to maximise their ability to understand a target through using the ear that has the best target-to-interferer ratio. For unilateral cochlear implantees better-ear listening will not be available, but they may still be able to make use of head shadow effects, by orientating their head so that their implanted ear is shadowed from the interferer and can attend to the target.

Figure 5.3 shows the predicted TIR polar response for a unilateral CI user, with the device implanted in their left ear. With the target located in front of the listener (red line), and the interferer rotated about them, there will be beneficial increases in TIR of up to 6.5 dB when the interferer is located on their non-implanted side, whereas when the interferer is to their implanted side the expected TIR will drop to levels below -3 dB. In comparison, when the interferer is located in front of the listener and the target rotated around them (blue line), the predicted TIR drops to as low as -8 dB when the target is to their non-implanted side, and reaches only 3.5 dB when the target is to their implanted side. In comparison, with both target and interferer rotated symmetrically around the listener (green line) beneficial TIRs of up to 10.5 dB can be achieved. This benefit of spatial separation is the same that is present for both normal hearing, and bilaterally implanted CI users, albeit at a reduced level. The difference between a unilateral and bilateral implantee is the range of situations for which the listener experiences a benefit. With bilateral CIs the benefit of symmetrical spatial separations are to both sides of the listener, whilst for unilateral implantees they occur only when the interferer is to the unimplanted side. For a right-ear implantee, with a target at 330° and an interferer at 30° significant gains in TIR would only be available to them by rotating 180°, and facing in completely the op-

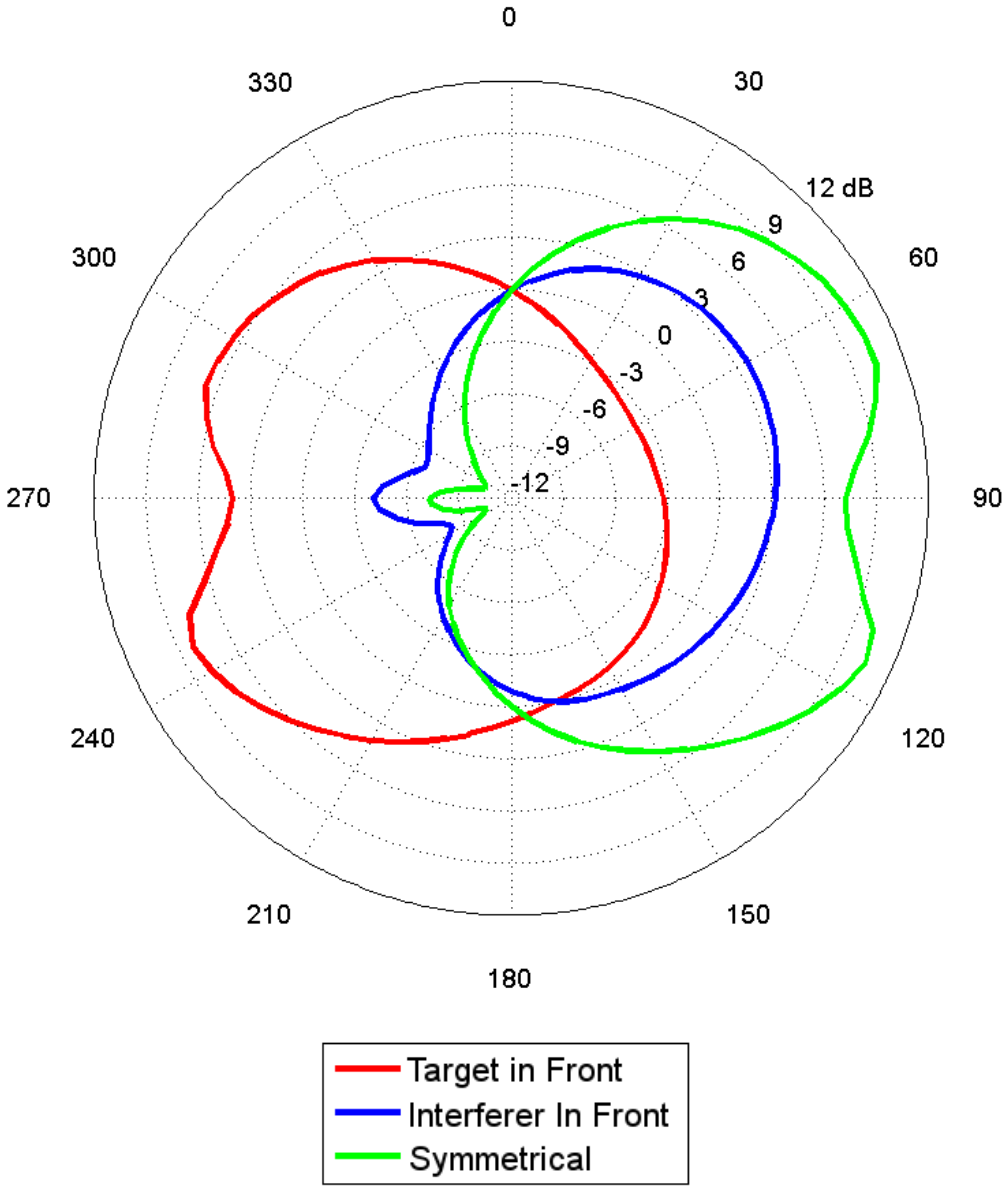


Figure 5.3: Polar plots of the modelled TIR with respect to source azimuth, either having a target in front at $0°$ and a moving interferer, an interferer in front at $0°$ and a moving target, or with neither source fixed, both moving around the listener in equal and opposite directions. Modelled as a unilateral CI user, device in their right ear.

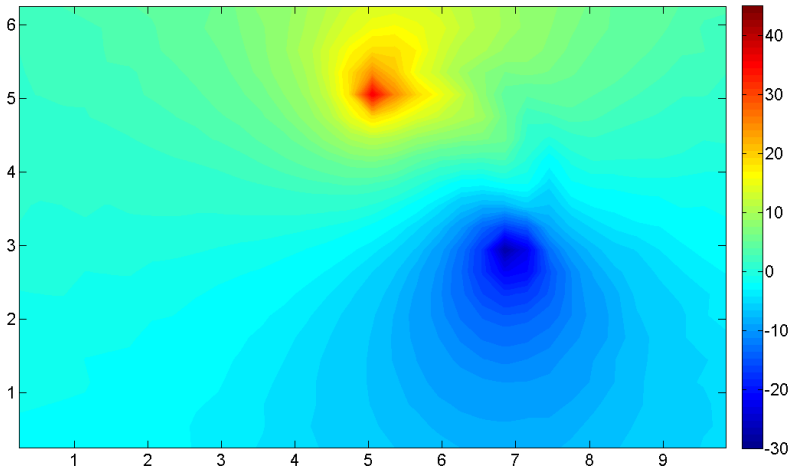
posite direction. This would be the same gain available to a bilateral implantee remaining stationary.

5.2 CI Room Maps

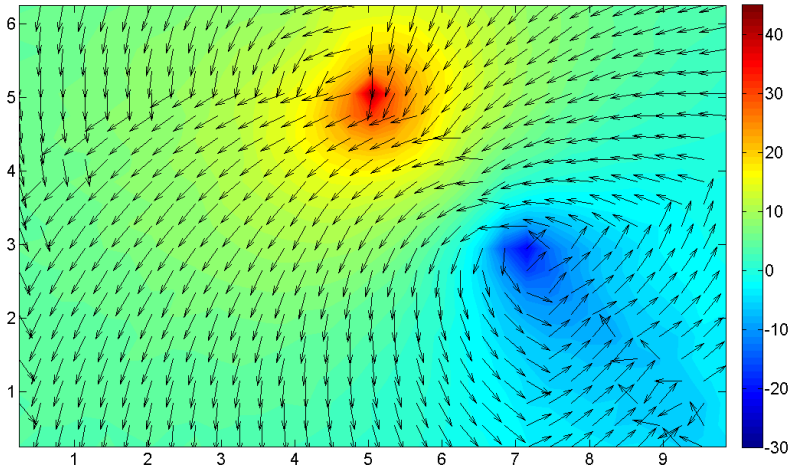
By modelling the listener as having no ability to use binaural unmasking, and either one or two ears, the effects and benefits of orientation on CI users within a room can be examined. For this modelling only a single interferer is used, using the anechoic "Int 3" configuration of the modelling carried out in chapter 4, (fig. 4.5). Figure 5.4 shows the results of this modelling for a unilateral CI user, implanted to their right ear, as (a) the expected TIR when the listener is facing the target, (b) the optimal direction and expected TIR when the listener is facing that direction, and (c) the benefit of orientation optimisation.

When the listener is facing the target the pattern of TIR produced is asymmetrical about the target-interferer axis. To the right of the axis the pattern of TIR is broadly the same as that produced for a normal hearing listener (fig. 4.9(a), p. 106), showing the same circular dip when the interferer is at $270°$. This is because at this point the implanted ear is the "shadowed" ear with respect to the interferer, and as such is likely to be the one used when the listener has the ability to use better-ear listening. To the left hand side of the axis, when the implanted ear is towards the interferer, the pattern is different, with a reduced level and no circular dip present.

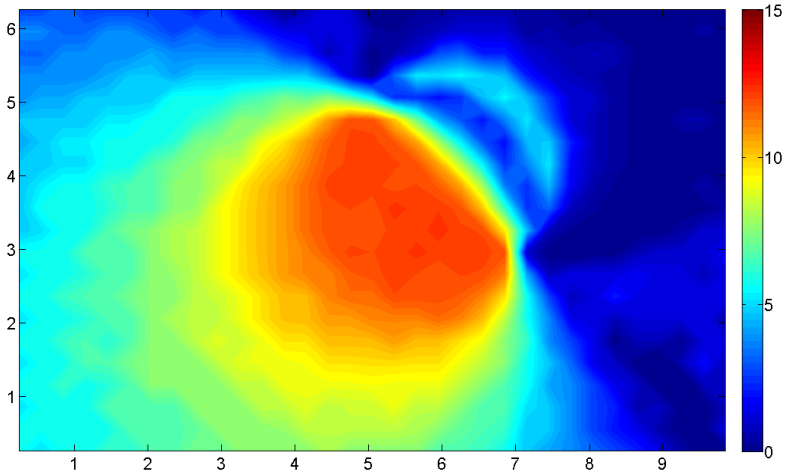
Once the effect of optimal head orientation is included, figure 5.4(b), it can be seen that for the majority of the locations the optimal orientation is where the interferer is located at approximately $270°$ to the listener, effectively directing their non-implanted ear towards the interfering source. This means that for the area to the right of the target-



(a)



(b)



(c)

Figure 5.4: Room maps, showing the modelled TIR when (a) listener is facing the target, (b) listener faces the optimum direction, and (c) the benefit of head rotation. Unilateral CI, implanted in the right ear. Target at 5 m by 5 m, Interferer at 7 m by 3 m.

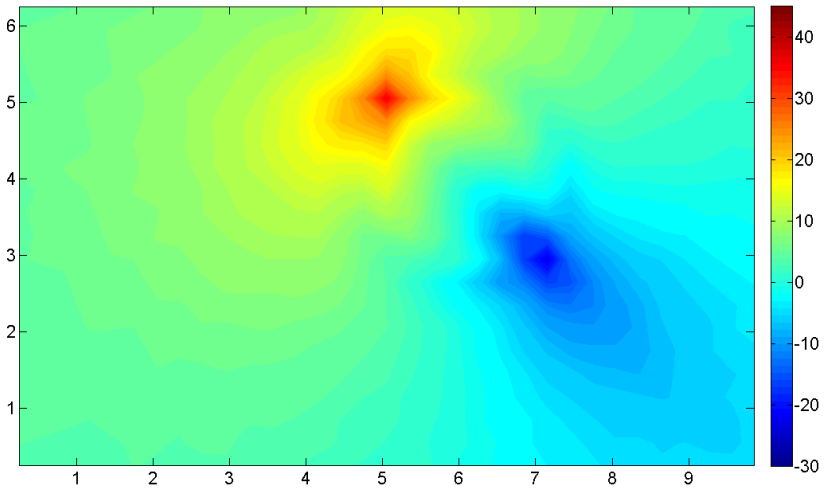
interferer axis the listener is facing the target, but for the much larger area to the left of the axis optimal orientation requires them to face completely away from the target, incurring the negative aspects discussed in Chapter 4.3.1. That said, the benefits available are great, with maximum values of 12.3 dB, and a mean value of 5.2 dB (s.d. is 3.8). Relevant statistics are contained within Appendix A, Tables A.6 and A.7.

The modelling was repeated, with the listener modelled as a bilateral CI user. This is, in effect, the same as modelling the listener as a normally hearing person, prior to the inclusion of the the benefit of binaural unmasking, and as such the pattern produced for when the listener is facing the target, figure 5.5(a), is the same as that produced in the earlier modelling (i.e. fig. 3.2(a), p. 58). The effect of implantation of the second ear is to raise the mean TIR by an average of 4.3 dB across all areas of the room (s.d.of 3.7).

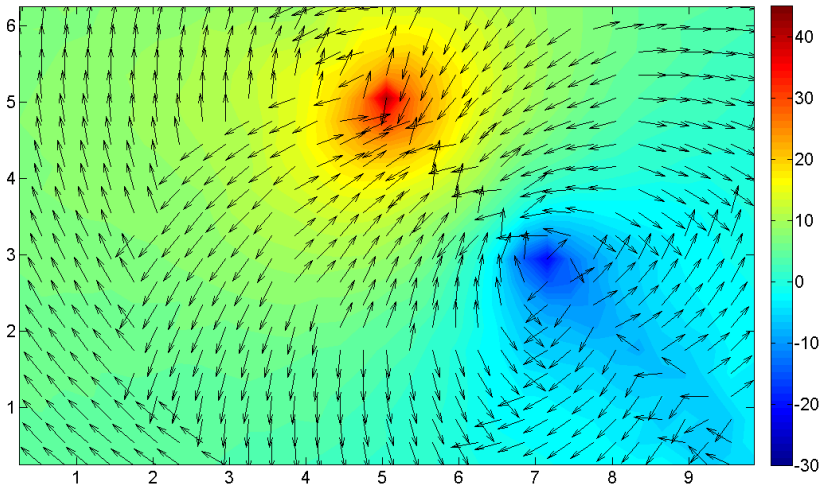
The effect of the second implantation on the orientation optimisation is that the TIR increases to the same level as observed in the unilateral implantation situation, but without the need for the listener to face directly away from the target in as many locations as previously observed. The benefits available from head orientation are reduced compared to unilateral implantation, but are achieved using only modest head movements that would not preclude lip reading.

5.3 Discussion

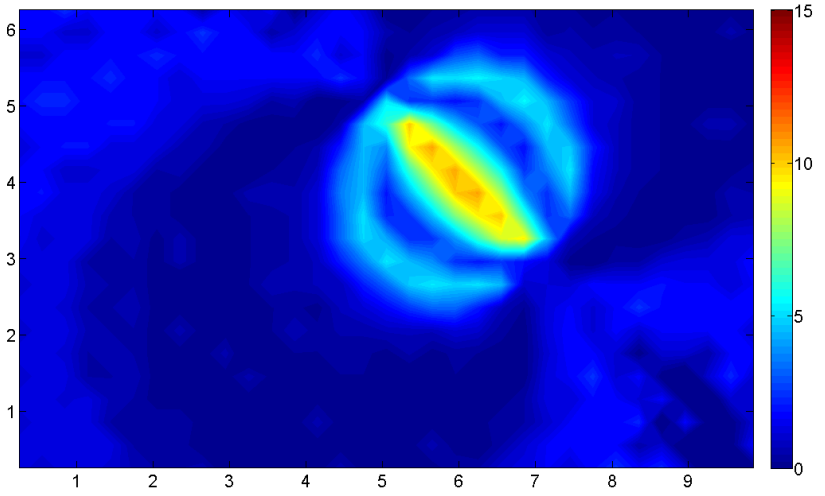
The National Institute for Health and Clinical Excellence, NICE, give guidance to the NHS in the UK about the cost-effectiveness of medical procedures based upon the cost of the procedure, or in this case the procedure plus the prosthetic device, against the improvement in Quality-Adjusted Life Years (QALYs). The QALY is a measure of the benefit a particular medical intervention has on the life of a patient. In assessing the



(a)



(b)



(c)

Figure 5.5: Room maps, showing TIR_{BE} when (a) listener is facing the target, (b) listener faces the optimum direction, and (c) the benefit of head rotation. Bilateral CI. Target at 5 m by 5 m, Interferer at 7 m by 3 m.

benefit of bilateral against unilateral implantation a technical appraisal was conducted on their behalf by Bond *et al* (2009). As part of this appraisal the authors analysed five studies with reference to the benefit of bilateral versus unilateral cochlear implantation. The NICE guidelines state that whilst the auditory outcomes were significantly better for bilateral implantation the results of speech perception were not, and as such the benefit gained was not great enough to be able to recommend bilateral implantation in adults.

As has been shown through the modelling carried out in this chapter, the situations that have been measured within the existing bilateral CI literature were not the optimal arrangements for demonstrating the effectiveness of bilateral implantation. Significant gains are available to CI users through head orientation, whether using a single device or a pair of them. But often for unilateral implantees these benefits are only found by turning right around.

Considering first the situation where a listener is facing the target, the difference in predicted TIR_{BE} between a listener with a unilateral CI and one with bilateral CIs is great. Differences up to 10.5 dB occur, dependent on listener location, and a mean difference of 4.3 dB (s.d. = 3.7 dB, see Appendix A - Table A.7). Figure 5.6(a) shows the pattern of this difference. As can be seen, there is a large area within the room where significant benefits are available to the listener just from having a second device, with no need to move or change their orientation.

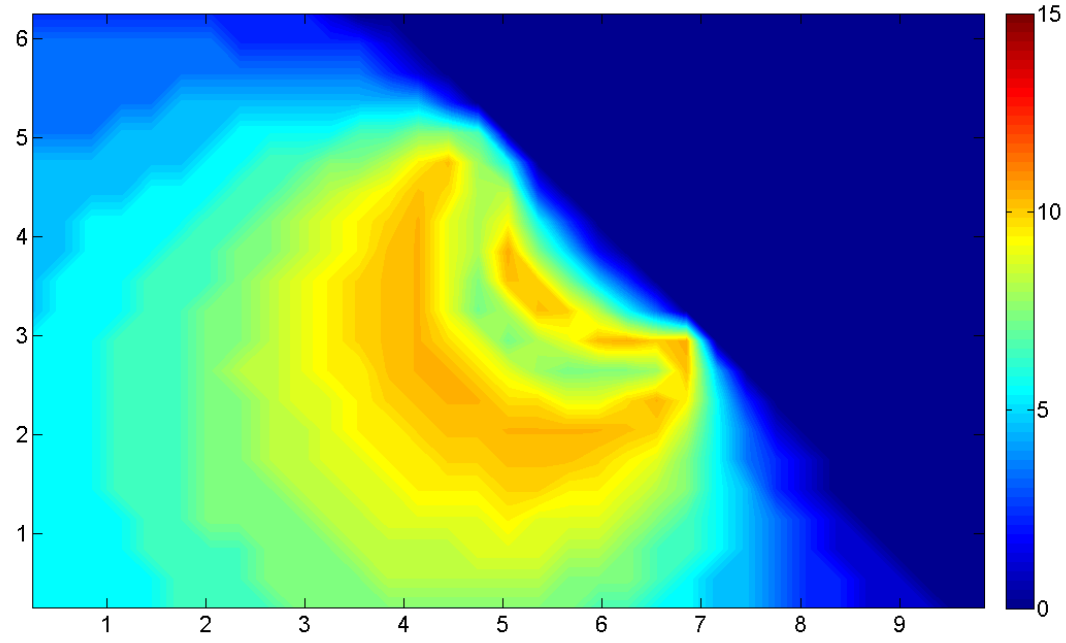
Now compare when the listener is orientated to their optimum direction. The difference in TIR_{BE} between bilateral and unilateral implantation is low, with a mean difference of 0.4 dB and a maximum difference of 1.5 dB, as shown in fig. 5.6(b). However, the key difference is the direction in which the listener needs to orient themselves to attain such a level. With unilateral implantation the listener often has to orient themselves so that they

are facing directly away from the speaker, removing the visual cues offered to comprehension, which are likely to be all the more important to a hearing impaired person, as well as going against all norms of social communication.

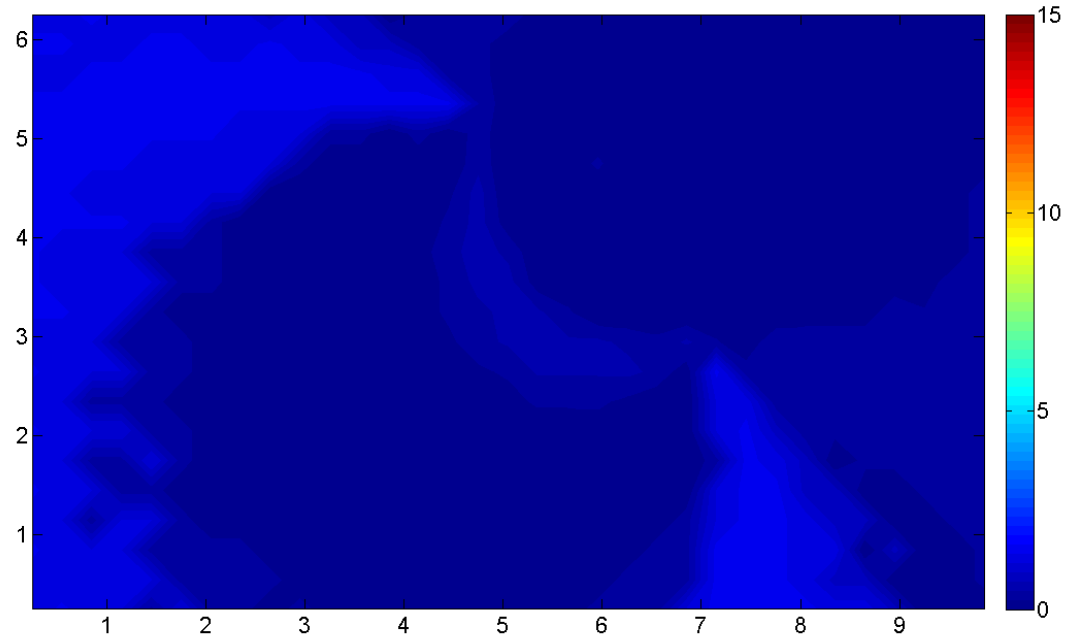
As with previous modelling, the inclusion of reverberation will be detrimental to the benefits afforded to CI users, as it is to normally hearing people. In the single-interferer modelling carried out in chapter 3, the effect of the addition of reverberation was to reduce the benefit of better ear listening by approximately 2 dB (from Table A.5), whilst also reducing the initial TIR before the inclusion of better-ear listening effects. Whether with unilateral or bilateral cochlear implants, reverberation will reduce the effects of head-shadow, reducing the effectiveness of optimal orientation and the benefits afforded to the listener.

There are a number of limitations to the modelling as carried out in this chapter, and how it relates to the real-world experiences of CI users. Primarily these are the discrete selection between better-ear listening effects and binaural unmasking, and the frequeny resolution of the model. With early CI devices that did not encode any of the temporal fine structure it is a clear decision that the CI user will receive little or no benefit from binaural unmasking, yet with advances in technology there may be greater possibility of the user having access to binaural unmasking, and as such the model will have to be altered accordingly. The model also assumes that the frequency resolution, and channel separation afforded to the listener with a CI is the same as that which is available to a normal hearing listener. With possibly only 16 or 24 channels available on the electrode array this level of resolution would not be possible. This could be modelled by adjusting the centre frequency and bandwidths of the gammatone filterbank in line with the frequency pattern as distributed along the array.

According to the implemented modelling the effect of bilateral implantation is great, allowing greater benefits to the listener than will be available to a unilateral CI user, though this finding would ideally need to be confirmed with empirical data.



(a)



(b)

Figure 5.6: Room maps, showing the benefits available to bilateral implantees over unilateral implantees, when (a) They are facing the target, and (b) they are facing the optimum orientation. Target at 5 m by 5 m, Interferer at 7 m by 3 m.

Chapter 6

Practical Room Configurations

In the previous room modelling carried out in this work, principally that in Chapters 3 and 4, there has been a maximum of 3 interfering sources, and the listener has often been distant to the target. In real life this is often not the case. Rather, the target and the listener are close together, with many interferers spread across the space. In this chapter it is this situation that is modelled, taking the example of a simplified restaurant situation, with target and listener sat at the same table, and interferers spread across the others.

The layout of the room used is shown in figure 6.1. It is a square 6.4 m by 6.4 m room, with a 2.5 m ceiling height. Nine tables are located across a 1.6 m square grid, each table is 50 cm square with two sources sat so as to be facing each other with a gap of 0.75 m between their heads, giving a total of 18 sources within the room. For all modelling in this chapter the tables are not present, so there is no reflections occurring from the surface of the table between the target and the listener.

The modelling was carried out by taking each table in turn as the test table, and, for that table, one source was said to be the target, and the other the listener. For the other eight tables, one of the two sources at each table was chosen to be an interferer. This

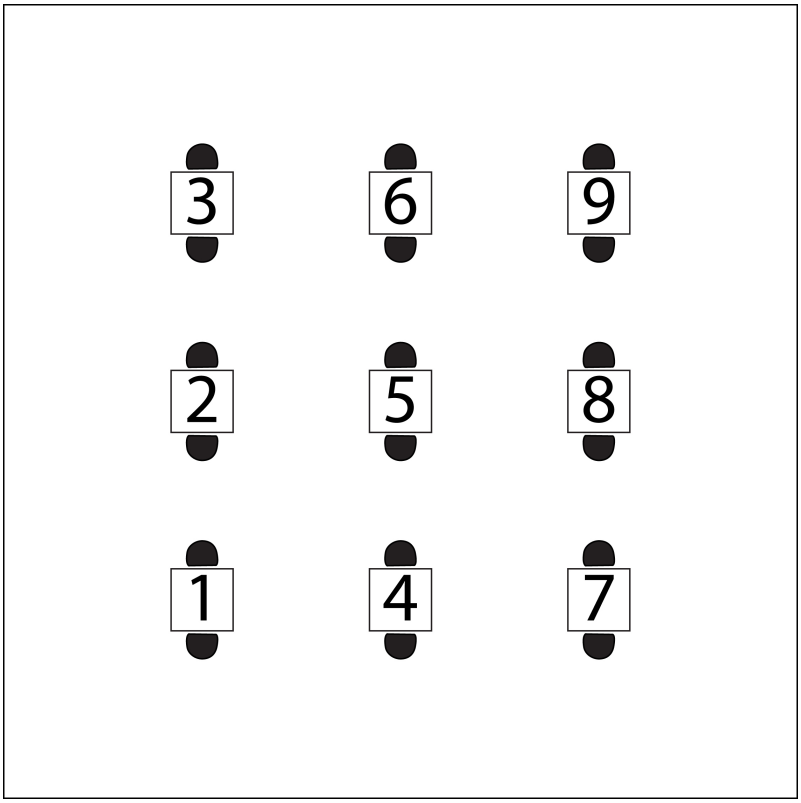


Figure 6.1: Diagram showing the layout of the tables within the 6.4 m by 6.4 m restaurant model, with tables centred on a 1.6 m by 1.6 m grid. Note - table surfaces are not present in the model, included here for indicative purposes only.

simulates a situation where there is a person trying to listen to the person across the table, with one person talking at each of the other tables in the room. For each iteration, a pseudorandom number generator, as implemented by the "rand" function in MatLab, was used to decide which of the two sources at the eight non-test tables was the interferer for each table. The modelling was carried out twenty times, randomly changing the interfering sources each time, before swapping the target and listener sources. This means that for each table forty cases were tested, 20 with one of the sources as the target, twenty with the other, and, for each case the interferers were randomly assigned. This was done so as to avoid modelling a situation in which all of the 16 interfering voices were talking simultaneously, which would not be realistic.

It was not the absolute levels of the noise that were of interest in this modelling, but rather it was the effect of table orientation and room acoustics on the ability of the listener to understand the target. A number of situations were therefore tested, starting first with a reference anechoic condition where all tables are aligned along the y-axis in an anechoic room as shown in 6.1. This initial case was then modified by first changing the alignment of some of the tables, before altering the room acoustics.

6.1 Reference Condition

The results of modelling the reference condition (listening through a single omnidirectional microphone in anechoic conditions) are shown in Figure 6.2 as either the omnidirectional TIR with no listener's head present (blue line), the TIR with better-ear listening effects (red line), or the TIR with both better-ear and binaural-unmasking effects (black line).

Consider first the omnidirectional TIR. Taking each table as the test table, the corner

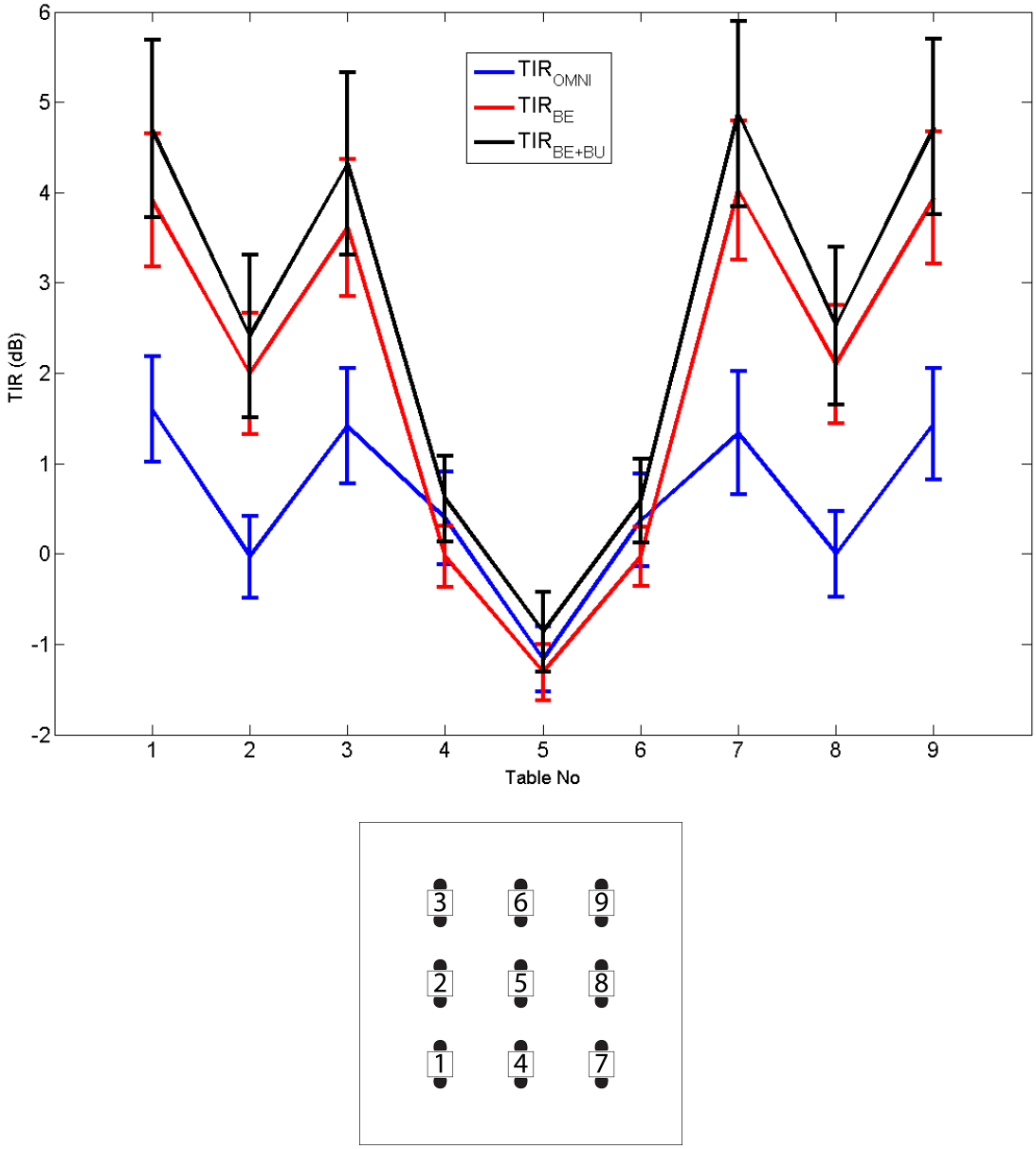


Figure 6.2: The results of modelling the anechoic restaurant configuration, with all tables facing the same direction. Top panel showing TIR with either an omnidirectional listener, listener with Better-Ear listening, or listener with both Better-Ear listening and Binaural Unmasking. Lower panel shows table arrangement.

tables, 1, 3, 7, and 9, give equal predicted TIR_{OMNI}s, whilst the tables in the centre of each side, tables 2, 4, 6, and 8, give a lower predicted TIR_{OMNI}. The lowest predicted TIR_{OMNI} is that for the centre table, 5. The reason the corner tables are best is that, in the omnidirectional case, the predicted TIR_{OMNI} is governed solely by the distance between the listener and the interfering sources. These tables have adjacent tables on two sides only, so most of the tables are at a greater distance. For the tables in the centre of each side the greatest possible average distance to the other tables is decreased, with competing tables on three sides. For the centre table, table 5, the average distance to the interferers is the lowest, with tables on all sides, and consequently the predicted TIR_{OMNI} is lowest for this case.

The effect of including the benefit of better-ear listening is to increase the predicted TIR for the listeners at tables in the left and right columns, 1, 2, 3 and 7, 8, and 9. The reason that these tables improve, and the centre column of tables do not, is that for the centre column, tables 4, 5, and 6, there are interfering sources to both the left and the right of the listener, and consequently better-ear listening is unable to help to any significant degree, whereas the left and right columns only have interferers to either the left or the right side of the listener, allowing better-ear listening to be more effective. Further benefits are observed for all tables when binaural unmasking is included, with an average increase of 0.63 dB.

6.2 Table Orientation

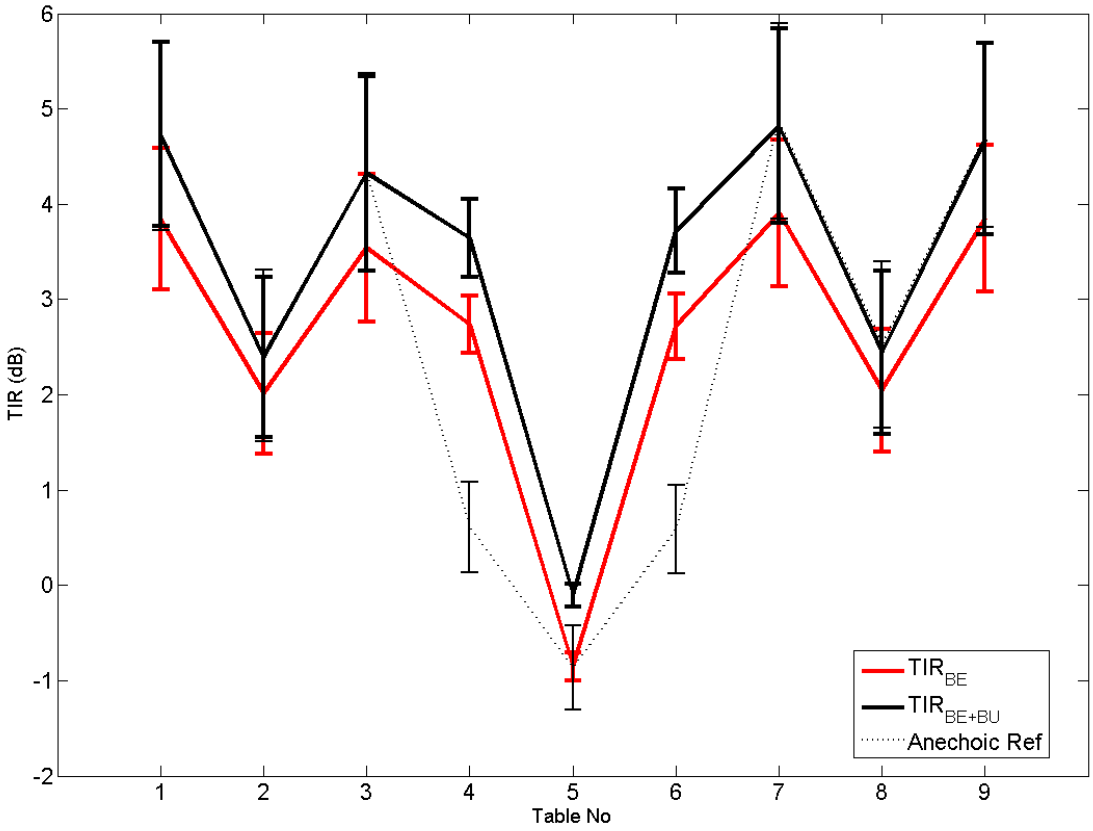
In many real-world restaurants, tables are not all aligned so as to be facing the same direction on a regular grid. Whilst retaining the regular grid positions of the tables it is possible to rotate them, to see the effects of rotation on the ability of the listener at that

table, as well as the effect on intelligibility at the other tables, aiming to maximise the ability of the listener at every table, whilst reducing the variation between tables. To look at this, two situations are tested. First, rotating tables 4 and 6 by $90°$, to see whether orientation wholly explains the difference between tables 2 and 8 and tables 4 and 6. Given that the average distance to the competing tables is the same, so the differences can not be explained by source proximetry, they may be explained by orientation. Secondly, freely rotating table 5, to see if there is any orientation that can help the listener significantly at the worst table.

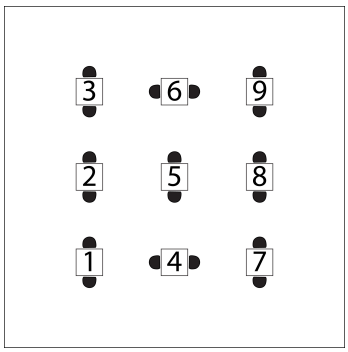
6.2.1 Rotating Tables 4 & 6

Given that the space is anechoic and symmetrical it is expected that tables 2 and 8 give the same predicted response, as their relative distances and orientations to the interfering sources within the room are equal. By rotating tables 4 and 6 by $90°$ they will be in a similar configuration as tables 2 and 8, with comparable interferer distances and orientations. As these two tables will become more similar to tables 2 and 8, an increase in the predicted TIR_{BE} and $\text{TIR}_{\text{BE+BU}}$ may be expected.

To test this, the model was run in the same fashion as for the reference anechoic condition, but with tables 4 and 6 rotated by $90°$, as shown in figure 6.3(b). The results of this analysis are shown in figure 6.3. In this instance the predicted TIR_{OMNI} is not shown, because it will be the same as that observed in the previous anechoic reference condition, but for comparison the predicted $\text{TIR}_{\text{BE+BU}}$ from the reference anechoic condition is reproduced as a dotted line. As can be seen, the rotated tables 4 and 6 give a higher TIR_{BE} and $\text{TIR}_{\text{BE+BU}}$ than before rotation by approximately 3 dB. This is due to the greater benefit of better-ear listening available to the listener now that one ear is orientated away from all



(a)



(b)

Figure 6.3: The results of modelling the anechoic restaurant configuration, with tables 4 and 6 rotated 90°. Top panel showing TIR with either a listener with Better-Ear listening, or a listener with both Better-Ear listening and Binaural Unmasking, as well as the anechoic reference condition of $\mathrm{TIR_{BE}}$ from Figure 6.2. Lower panel shows table arrangement.

of the interferers, and towards an anechoic wall. Tables 4 and 6 now give predicted levels of intelligibility greater than at tables 2 and 8. This is likely due to the fact that unlike at tables 2 and 8 the interferers at the tables on the target-listener axis are not directly in front or behind the listener, but are slightly offset to one side or the other, allowing for an increased use of spatial unmasking.

Rotation of tables 4 and 6 does not affect the predicted response at the other tables, except for the centre table, which sees a small improvement. The latter effect is probably due to the fact that the interferers at tables 4 and 6 are no longer directly in front or behind of the listener, on the same axis as the target, but are now on a different azimuth.

6.2.2 Rotating Table 5

Using the situation where tables 4 and 6 are rotated as the starting configuration, the effect of rotating the centre table, table 5, can be examined. In this instance only the response of table 5 is plotted as a polar diagram, as shown in figure 6.4, with 0° as the original orientation, showing only the $\text{TIR}_{\text{BE+BU}}$. From the polar plot it can be seen that small changes in orientation away from the starting position can have large effects on the predicted $\text{TIR}_{\text{BE+BU}}$, a rotation of 20° gives an increase in $\text{TIR}_{\text{BE+BU}}$ of 3 dB, with a rotation of 60° giving the greatest increase of approximately 4.2 dB. At 60° there will be no interfering sources on the target-listener axis, allowing for the greatest effective use of better-ear listening and binaural unmasking as explored previously (Chapter 4).

6.3 Room Acoustics

Anechoic rooms are not commonly found outside of industrial and academic research environments, and the chances of having an anechoic restaurant are slim. So the effect

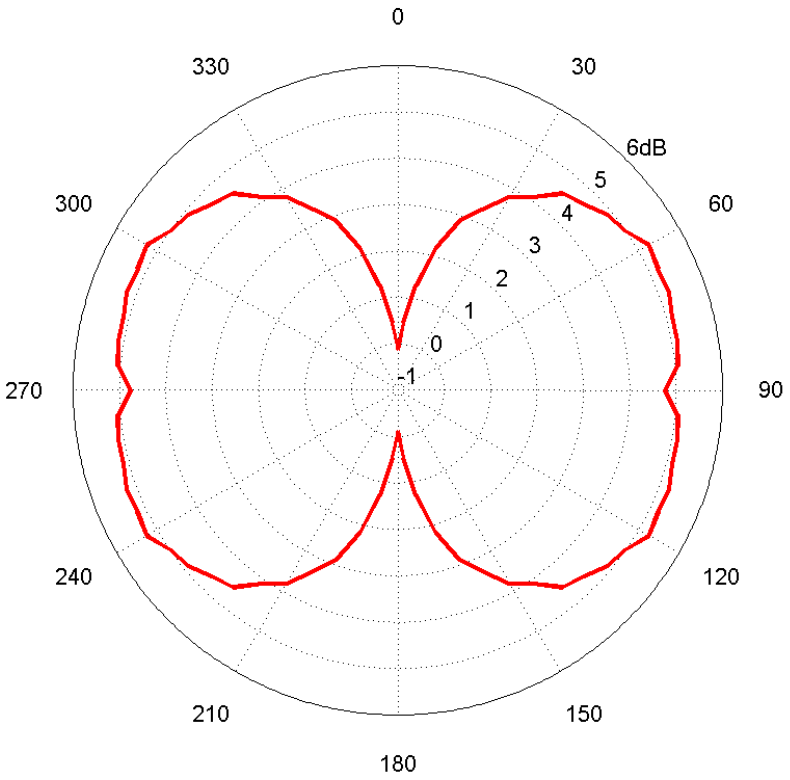


Figure 6.4: The results of modelling the anechoic restaurant configuration, rotating the centre table, showing the average TIR_{BE+BU} for the two listeners relative to table orientation. One interferer present for each of the other 8 tables.

of reverberation in a multiple-interferer complex situation is of interest. To look at this, two configurations of room acoustics were analysed, with different placements of acoustic treatment, either an acoustically absorptive ceiling, or absorptive walls. In both instances, the absorption coefficients were chosen so as to keep the overall statistical reverberation time of the room the same, allowing for differences in the direction of the reflections, but keeping the overall level of reflections constant. Using the Sabine equation (Eq. 3.1, p. 55), the RT_{60} of the rooms is 0.33 seconds at all frequencies.

6.3.1 Absorptive Ceiling

To model a room with an absorptive acoustic ceiling the absorption coefficients of the surfaces were set to 0.05 for the walls, 0.07 for the floor, and 0.9 for the ceiling. These are broadband approximations of painted plaster on masonry walls, and wooden block flooring on a concrete floor. The same modelling was carried out as for the anechoic condition, firstly with all tables facing the same direction (figure 6.5), with tables 4 and 6 rotated, (figure 6.7) and finally rotating the centre table (figure 6.8). Note that for figures 6.5 and 6.7 the dB scale is reduced compared to figures 6.2 and 6.3.

In reverberation the predicted TIR_{OMNI} when all tables are facing the same direction (Blue line, figure 6.5) is equal at all tables except table 5, which unlike in the previous anechoic condition performs better than the other tables. This is likely to be due to the fact that at all other tables strong reflections will occur from the walls nearest to the listeners, which are reduced in level by the time they reach the centre table. Once the head is included, and the benefit of better-ear listening that it affords, the tables in the left and right columns again outperform the centre column, though interestingly the centre table performs worse when the listeners have heads than when they do not. The cause of this

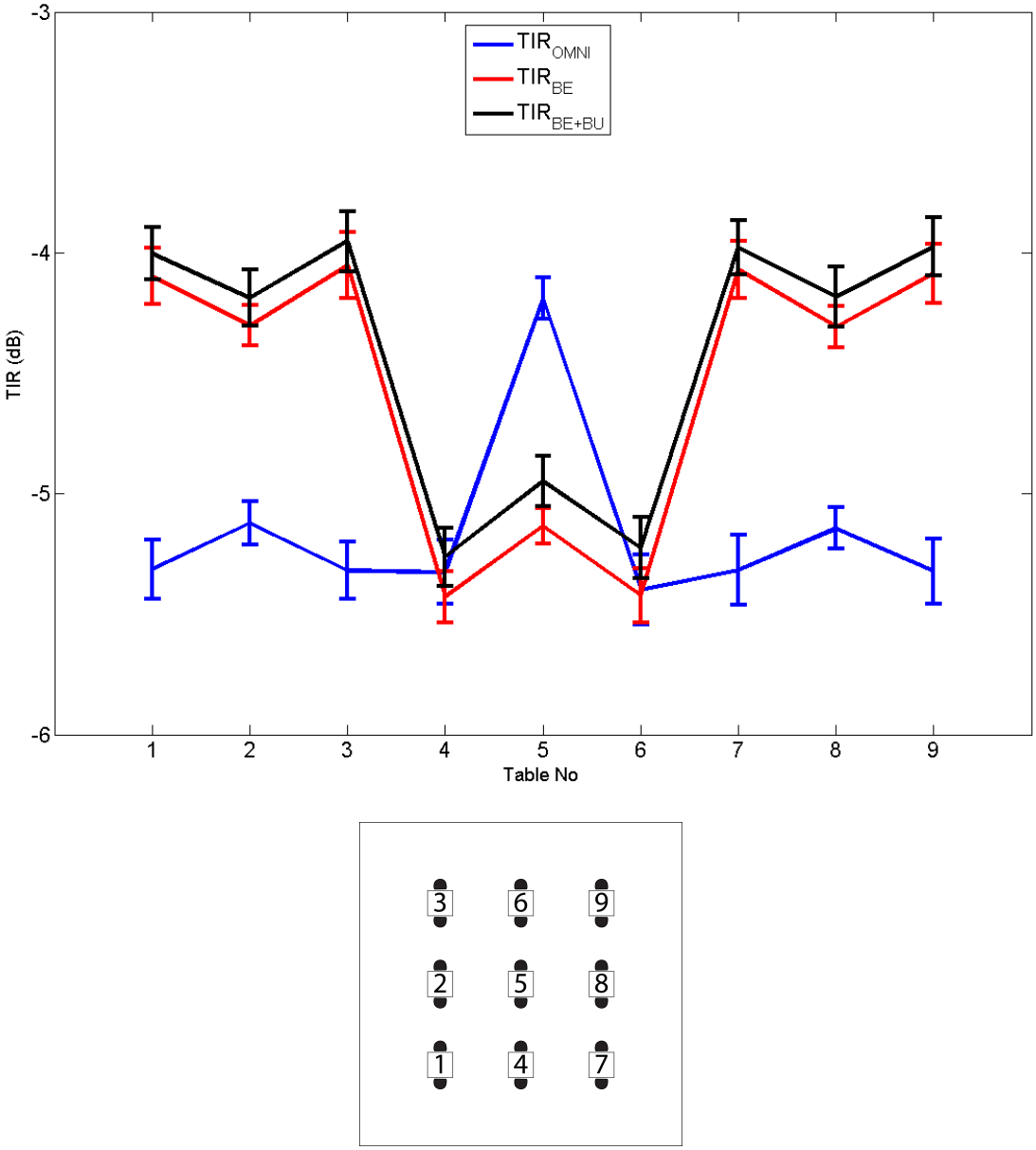
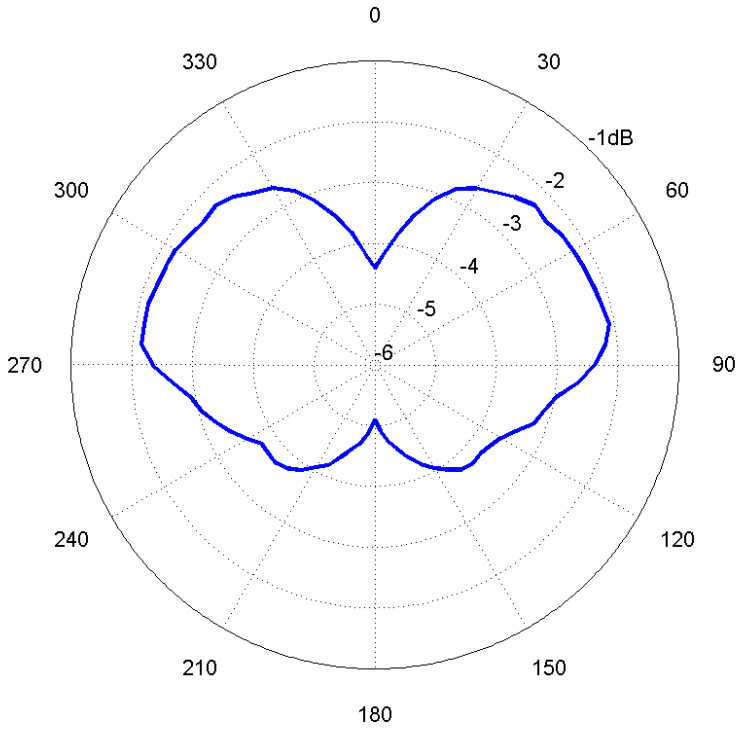


Figure 6.5: The results of modelling the reverberant restaurant configuration, with an absorptive ceiling, and all tables facing the same direction. Top panel showing TIR with either an omnidirectional listener, listener with Better-Ear listening, or listener with both Better-Ear listening and Binaural Unmasking. Lower panel shows table arrangement.

(a)



(b)

Figure 6.6: (a) The directivity of the left and right ears of a KEMAR manikin, with SII speech weighted frequency selectivity, and (b) the effect of rotating a single listener at table 5, independently of the table, showing $\mathrm{TIR}_{\mathrm{BE}}$ with respect to azimuth.

is the directional response of the listener. The polar response of the ears is such that a source directly in front of the listener is not picked up as well as one at $\pm 45^{\circ}$. This means that the level of target received by the omnidirectional microphone used to model the $\mathrm{TIR}_{\mathrm{OMNI}}$ is greater than that received at the ears when directly facing the target, as they are effectively shadowed by the head. Because of this, head rotations will be of benefit to the listener, through placing the target in the more sensitive region of one of the ears. Figure 6.6(a) shows the polar response of the two ears of the KEMAR manikin, with the lower level of response in both ears at 0° than at $\pm 45^{\circ}$. Figure 6.6(b) shows the predicted $\mathrm{TIR}_{\mathrm{BE+BU}}$ that occurs by rotating the listener at table 5 independently of the table. This shows that rotating so that the target is to one side of the listener rather than directly in front increases performance. Binaural unmasking benefits the listener at all tables, though to a lesser extent, on average less than 0.2 dB.

Rotation of tables 4 and 6, as shown in figure 6.7, increases the predicted $\mathrm{TIR}_{\mathrm{BE}}$ at these tables, and increases the benefit of binaural unmasking, not only for tables 4 and 6, but also for table 5, due to no longer having sources on the target-listener axis.

Rotating the centre table produces the same pattern of benefit as that produced in the anechoic condition, albeit at a reduced level, as shown in figure 6.8, with the maximum orientation being at 60° and a relative benefit of 4.6 dB. Whilst the addition of reverberation for a room with an acoustic ceiling has reduced the absolute level predicted at the centre table compared to the anechoic equivalent (figure 6.4), the relative increases due to rotation of the table remain approximately equal.

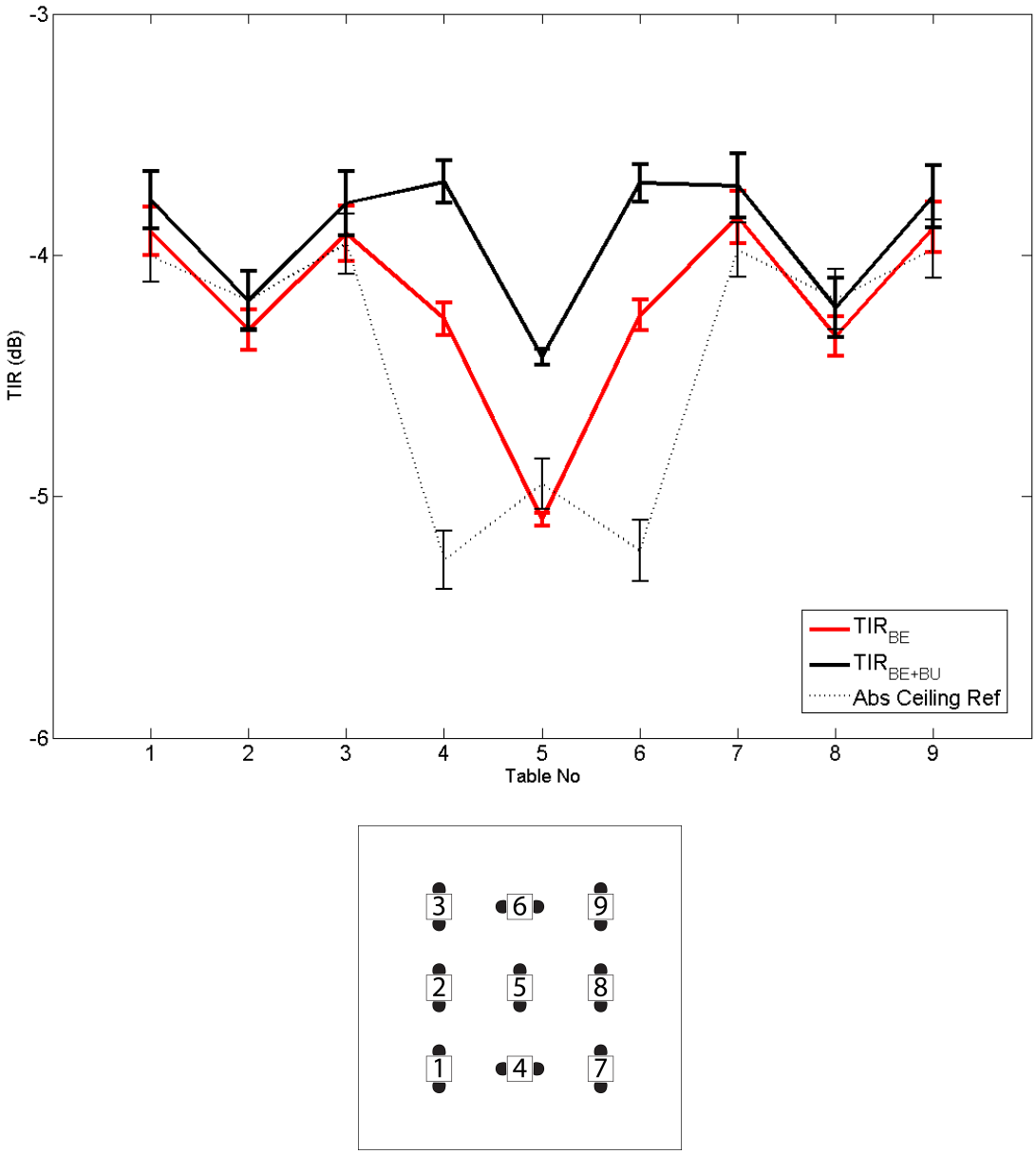


Figure 6.7: The results of modelling the reverberant restaurant configuration, with an absorptive ceiling, and with tables 4 and 6 rotated 90°. Top panel showing TIR with either a listener with Better-Ear listening, or a listener with both Better-Ear listening and Binaural Unmasking, as well as the absorptive ceiling reference condition of TIR_{BE} from Figure 6.5. Lower panel shows table arrangement.

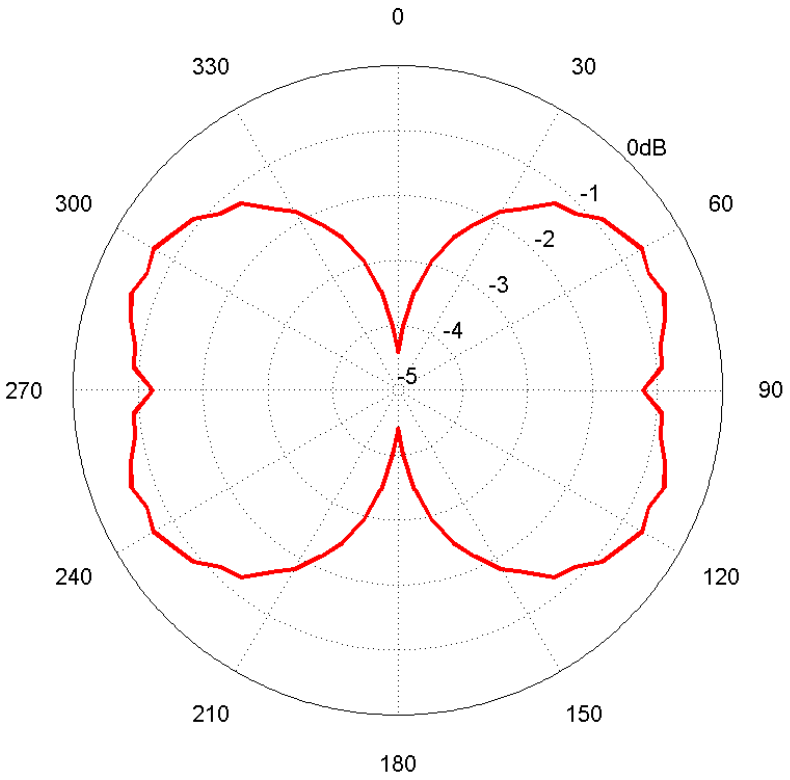


Figure 6.8: The results of modelling the reverberant restaurant configuration, with an absorptive ceiling, and rotating the centre table. Showing the average TIR_{BE+BU} for the two listeners with respect to azimuth.

6.3.2 Absorptive Walls

The other modelled configuration of room absorption is that in which the walls are acoustically absorbent and the ceiling is reflective. To model this the absorption coefficients of the room were changed so that all of the walls had a broadband value of 0.6, with a ceiling of 0.05 and floor of 0.07. This maintains the reverberation time of 0.33 seconds whilst reducing the intensity of lateral reflections.

The first effect of this is that when modelling the omnidirectional response of the listener, all tables are equal, with table 5 no longer doing significantly better than any other table (fig. 6.9), as it did with the absorptive ceiling (fig. 6.5). Given that the main difference between the centre table and the outer tables in the absorbent ceiling condition was the proximity to the reflective surface, this is now reduced, with all tables having the same distance between them and the reflective ceiling and floor. The benefit due to better-ear listening is greater with absorptive walls than with an absorptive ceiling, with a mean benefit of 1.1 dB compared to 0.7 dB, though there is no significant benefit due to binaural unmasking.

Rotation of tables 4 and 6 (figure 6.10), again gives an increased TIR_{BE} for these tables, and also increases the benefit of binaural-unmasking at both tables 4 and 6, as well as the centre table, table 5. As in the anechoic case, the increased TIR_{BE+BU} at tables 4 and 6 is likely to be because one ear is now facing a relatively absorbent surface, away from all of the interfering sources.

The rotation of the centre table has the same outcome as in both the anechoic and the absorbent ceiling conditions, with both the same pattern and range of benefits. This is probably due to the orientation of the sources being more important to the performance of the listener than the acoustic environment, at least in these situations where the distance

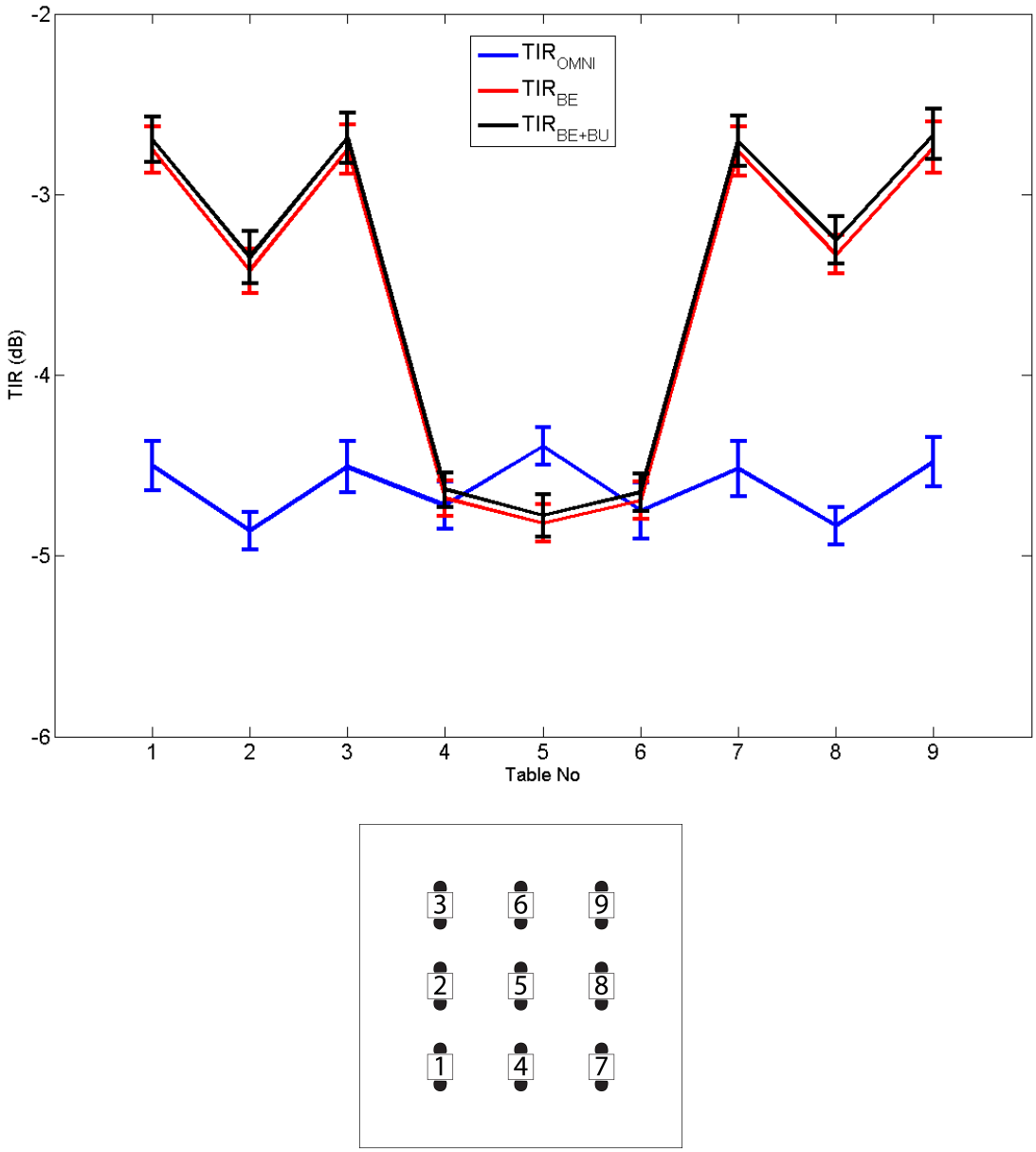


Figure 6.9: The results of modelling the reverberant restaurant configuration, with absorptive walls, and all tables facing the same direction. Top panel showing TIR with either an omnidirectional listener, listener with Better-Ear listening, or listener with both Better-Ear listening and Binaural Unmasking. Lower panel shows table arrangement.

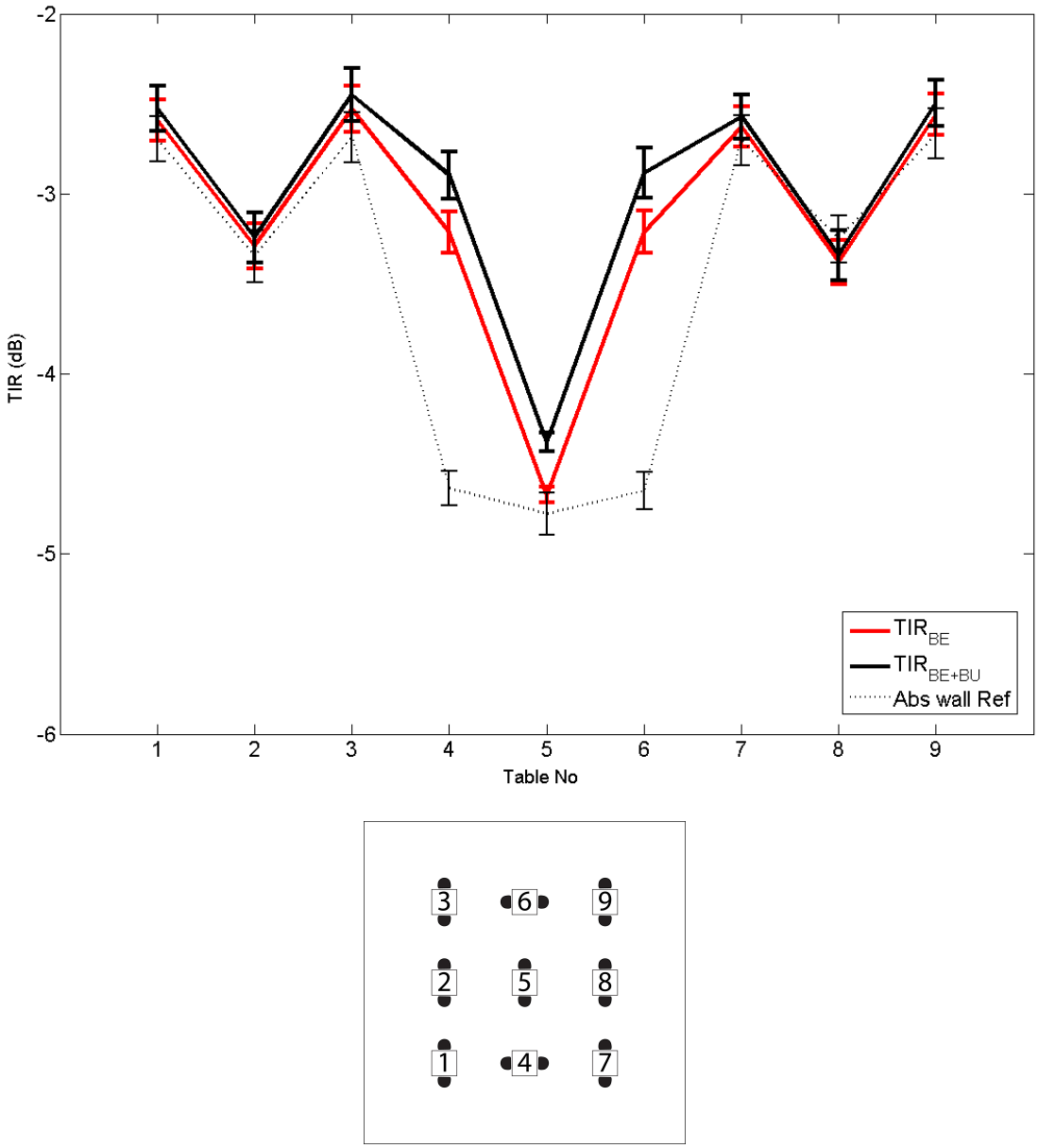


Figure 6.10: The results of modelling the reverberant restaurant configuration, with absorptive walls, and with tables 4 and 6 rotated 90°. Top panel showing TIR with either a listener with Better-Ear listening, or a listener with both Better-Ear listening and Binaural Unmasking, as well as the absorptive wall reference condition of TIR_{BE} from Figure 6.9. Lower panel shows table arrangement.

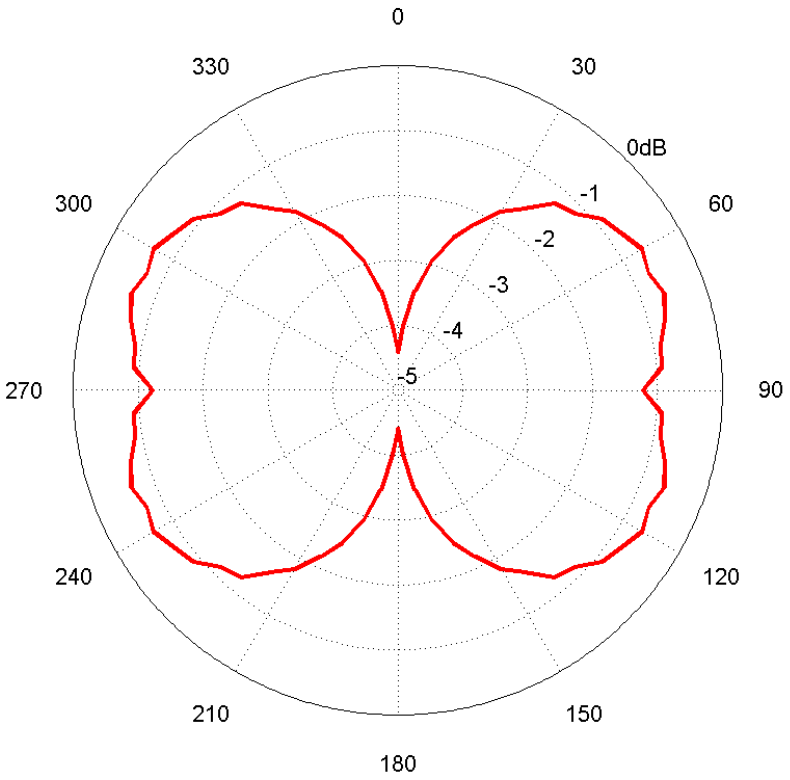


Figure 6.11: The results of modelling the reverberant restaurant configuration, with absorptive walls, and rotating the centre table. Showing the average predicted TIR_{BE+BU} for the two listeners with respect to azimuth.

between listener and interferer is relatively low, with a consequently high level of direct sound.

6.4 Discussion

Restaurants pose a much more complex environment for a listener than is typically studied experimentally. Whilst previous modelling has shown the effects of head orientation and the effect of reverberation on speech understanding in quasi-experimental situations, the modelling of a restaurant gives a view of a differing situation.

There are two key differences between this modelling and that which was done previously, namely that the target and the listener were closer to one another than to the interferers, which are now greater in number, and that when using optimal orientations both target and listener were moved as a pair, compared to the previous modelling where only the listener was moved.

Given the number of interferers, and the proximity of the target and listener to them, it is not surprising that the benefits available to the listener through the use of better-ear listening and binaural unmasking are reduced when compared to the previous modelling. In the anechoic condition, the maximum combined benefit, B_{BE+BU}, was 3.5 dB, whereas in the single-interferer anechoic modelling carried out in chapter 3, maximum combined benefits exceeded 10 dB. This is a significant reduction in benefit.

As expected, given previous modelling, the general effect of the addition of reverberation is to reduce both the initial TIR_{OMNI} for each table, as well as reducing the magnitude of the binaural benefits available to the listener. Reverberation occurring in a room with acoustically absorptive walls gives a greater TIR_{OMNI} and larger benefits due to better-ear listening, particularly at tables 2 and 8, compared to a room with an absorptive ceiling.

We can see this effect through the reduction of lateral reflections reaching the listener, which will have different ITDs and ILDs compared to the direct sound. As shown in the previous modelling, such reflections will impair the ability of the listener to use both better-ear listening and binaural unmasking. One limitation of having carried out the modelling in a square room is that the geometry of the space may have caused high levels of model reflections at certain frequencies that would be particularly prevalent at the centre of the space, which may have unduly affected the results as predicted throughout for table 5. Careful choice of the room geometry would minimise these room modes, whilst fitting within the constraints of having a rectangular regular room. A number of studies have suggested room geometry ratios that will produce even spaced room modes, stopping them affecting one frequency range greater than others, such as the 2:3:5 ratio of Volkmann (1942), or the 1:1.26:1.59 ratio of Boner (1942). Sepmeyer (1965) and Louden (1971) give several other favourable room ratios to choose from.

Table orientation will play a major part in the ability of a listener to understand a target at the same table. In this modelling only three of the nine tables were rotated, and then only one of them was freely rotated to find the optimum orientation. Taking this into account, large benefits were seen when rotating tables 4 and 6 so that the persons are facing along the wall nearest to them, rather than toward or away from it. This will be due to shadowing of the ear closest to the wall with respect to the interferers, and may well be a general effect for any table against a wall, irrespective of room geometry. Further modelling will be needed to ascertain whether there are greater benefits with rotations other than 90° for tables adjacent to a wall. That said, whilst greater benefits may be available, a requirement for the restaurant to have all of the tables set at odd angles will likely make it highly impractical.

Rotations of the centre table allow the listener benefits approaching 5 dB in both the anechoic and reverberant conditions. It would similarly be possible to model the room allowing each table free rotation to decide on the optimum orientation of each table, whilst remaining centred on the regular grid. Through doing this it would be possible to find the combination of table orientations that give the best overall level of intelligibility at all tables, and the lowest variation across tables. Though if the tables are limited to minimum rotations of 5° there will be 72^9 possible combinations of table orientations. Whilst possible, even with the revised model this will be a time consuming task. That said, it is possible to gain some general guides from this modelling, at least for this small sample of room configurations. Principally that tables that are near to a wall need to be orientated so that the target and listeners are facing along the wall, rather than perpendicular to it, and that tables in the middle of the room are better when orientated so that the target-listener axis has no interfering sources positioned along it.

Whilst the modelling carried out in this chapter is limited in scope, it illustrates a principle that can be extended to encompass other configurations, with greater numbers of sources per table and table orientations, as well as larger numbers of tables. Tackling more complex cases is merely a question of computation time.

Chapter 7

General Discussion

In the introduction to this thesis a number of aims were set out for the research project. Principally, they were to develop an accurate model for predicting the intelligibility of speech in noise, capable of taking into account the effects of binaural hearing, and to use this model to investigate some of the ways in which both the position and number of interferers affects speech intelligibility, as well as the effect that the acoustic design of rooms has on speech intelligibility. It was also planned to look at the effect that listener orientation, as well as listener position in a space, has on their ability to understand a target voice. In attempting to address these aims, a model was developed capable of predicting target-to-interferer ratios that embodies the benefits of better-ear listening and binaural unmasking, whilst being able to handle multiple interferers, in any acoustic environment.

This general discussion is divided into three sections, looking first at the model proposed in Chapter 2, what key benefits that it gives over previous models, as well as the limitations of the model. The second section looks at the results of the ways in which the model has been applied throughout this thesis, and the general principles that can be taken from the situations tested. In the final section, ways in which the model can

be developed further are discussed, including greater analysis of architectural design and possible applications within the field of audiology, as well as audio reproduction.

7.1 The Revised Model

7.1.1 Speed & Accuracy

Lavandier & Culling (2010), proposed a method to predict Speech Reception Threshold (SRT) measurements for speech in noise. Whilst their model was accurate for their data set it was felt that improvements could be made, allowing the method to be of greater use. As explained in chapter 2 the principal way in which the revised model and the original model of Lavandier & Culling differs is in the fact that the original model works on comparably long acoustic noise recordings convolved with impulse responses, whilst the revised model works directly on the measured impulse responses. This change benefits both speed and accuracy, as well as reducing the amount of data that needs to be processed and stored in order to carry out the calculations.

The increase in speed is due not only to the change in input to the model from convolved noise to binaural room impulse responses, but as well as the removal of the need to calculate the cochlear excitation patterns in calculating the benefit of better-ear listening.

To attain the same level of prediction accuracy the noise recordings used in the model of Lavandier & Culling need to be significantly longer than the revised model impulse responses, by a factor of 100, as described in Chapter 2.1.2. This is due to the stochastic nature of the noise sources, and the difficulty in extracting an accurate coherence value that this causes.

This greatly reduces the amount of data that needs to be processed by the model. Tak-

ing as an example the complex optimum orientation maps generated in Chapter 4.2, in each map there are 693 listening locations, each one needing 72 predictions to find the optimum head orientation for that location. This results in 49,896 individual configurations to be modelled. If for each calculation the time taken is decreased by 1 second, an overall saving of nearly 14 hours is attained. In this instance the impulse responses used to generate the map were 10,000 points long, with 49,869 configurations, up to 4 sources per configuration, and 2 impulses per source, giving a total of approximately 7.6 GB of raw data to be stored. If this was all to be convolved with long noise samples for computation with the original model of Lavandier & Culling there would be over three quarters of a terabyte of data to be stored.

It is difficult to directly compare the speed of the original model of Lavandier & Culling to the revised model due to the differences in the way each model has been implemented. In the original work of Lavandier & Culling the model was implemented using |WAVE (Culling 1996) on a Unix workstation, whilst the revised model has been implemented using MatLab on a Microsoft Windows workstation, both machines running different hardware. Even taking these differences into account the comparative speed of the revised model is great. To predict the 16 conditions used in Lavandier & Culling (2010) with the original model took the authors slightly over 2 hours of modelling time (7 mins 40 seconds per configuration), whilst the revised model is able to predict the same configurations to the same level of accuracy in 40 seconds (2.7 seconds per configuration). Using the same implementation of the model as Lavandier & Culling to generate one of the optimum orientation maps in Chapter 4.2 would take approximately 265 days, rather than the 37.5 hours taken by the revised model.

In many areas of modelling, as well as the wider world, speed comes as a trade off

against accuracy. In the case of the revised model, however, greater accuracy comes at the same time as increased speed. By using impulse responses rather than noise recordings the stochastic error inherent in the noise is removed. The accuracy of the model is reliant upon the quality of its input. If the input accurately reproduces what will be heard by the listener, as do the impulse responses used in the validation of the model (chapter 2), the model will be able to accurately predict the level of intelligibility.

7.1.2 Limitations

Whilst the model is fast and highly accurate, it has to be acknowledged that there are a number of limitations in what the model can presently handle, chiefly relating to the type of interferer used, and the effect of reverberation on the target speech, but also the fact that it predicts relative speech intelligibility levels, not absolute levels.

7.1.2.1 Relative Speech Intelligibility

The model predicts a target-to-interferer ratio at the threshold of intelligibility (when listeners can understand 50% of the spoken words). This is a relative measure of speech intelligibility, and in order to predict a given speech reception threshold it requires a known reference condition. It may be possible to use a method similar to that used in the computation of the Speech Intelligibility Index (ANSI 1997) to convert from the target-to-interferer ratio (the equivalent of the signal-to-noise ratio used in the SII) to an absolute index of intelligibility. Further investigation would be needed to confirm the feasibility of this, and as such the model currently can only predict a relative intelligibility measure. Also note, that for greater levels of speech understanding above threshold, at say 80 or 90% of words understood correctly the relative benefits of binaural unmasking and better-

ear listening will be different, and the model will not currenlty predict those correctly.

7.1.2.2 Interferer Type

One of the principal limits of the present model is that it cannot take account of the type of interferer. In all of the modelling in this thesis, the interferer has always been assumed to be a continuous noise interferer, such as an air handling system. With a more complex interferer, particularly speech, there are a number of other factors that will either help or hinder the listener and which the model cannot yet handle. Natural speech is a complex signal, with a changing frequency pattern, varying level segments of periodic sound, and intermittent gaps.

The sonorant portions of speech, such as vowels and nasals have a fundamental frequency, F_0, which changes with time. Each individual person's voice has a different mean F_0, defined by their physiology. Differences in F_0 can allow for a listener to segregate competing voices (Brokx & Nooteboom 1982). Even when two voices have the same mean F_0 instantaneous differences will allow the listener to segregate the voices.

Another effect of a real speech interferer is that the level of real speech will vary over time, with the talker even becoming silent occasionally. This allows the listener to hear the target better in both the gaps between words of the masking speech, but also in the quieter portions of the masker words. As such the masker, as predicted by the current model, will vary over time (Rhebergen & Versfeld 2005). The model currently has no mechanism for predicting the overall effect of these variations.

In addition to these factors which influence "energetic masking", there are also "informational masking" effects which can become prominent with speech interferers, For instance, the words of two concurrent sentences may become confused. Such confusions

between competing voices may be overcome by a difference in mean F_0 or in perceived sound location. Alternatively the listener may segregate two competing voices by the subject they are speaking about. Normal conversation follows a socially accepted set of rules, and one of the effects of this is that changes in topic are either slow or heavily defined. As a listener we know the subject of discourse of the attended voice, and can use this to help separate out other voices talking on differing subjects.

The field of speech-on-speech segregation is a hot topic in current research. Hopefully as these mechanisms are better understood more of these factors can be combined into the model.

7.1.2.3 Target Reverberation

As part of the model validation, a number of data sets from the literature were modelled, as well as unpublished data collected in the laboratory. In the data from the literature, either all of the sources were anechoic, or the target was anechoic and the interferer reverberant (in the case of Lavandier & Culling (2010)). The only data set modelled that has a reverberant target is the empirical data collected in the laboratory (see 2.2.6, p. 41). In all of the conditions, the target was close to the listener, and therefore had a high direct-to-reverberant ratio.

Because of the way the model presently works, using all of the target impulse response to calculate the better-ear listening target-to-interferer ratio, all energy contained within the target is considered to be of use to the listener. Whilst this assumption has proven accurate with targets that have a high direct-to-reverberant ratio, it may be less accurate when the direct-to-reverberant ratio decreases. Losses in accuracy may be expected due to the temporal smearing of the speech that will occur in reverberation, making it difficult

to understand what the target voice is saying, either in the presence of an interferer or in quiet. As a result the model will predict better performance than observed empirically.

7.1.3 Revised Model Conclusions

The model does not in itself tell us anything new about hearing or about architecture, but rather it gives us a tool that enables us to look at things in a different way. Previous models of speech intelligibility either did not include the fact that human listeners have two ears (SII, STI), or their processing was very slow. In comparison to the revised model of Beutelmann *et al* (2006, 2010), the closest published model in its design to the model developed here, this model maintains a lower level of processor overheads, and does not require acoustic waveforms to be able to calculate binaural speech intelligibility. Beutelmann *et al* were also only able to predict SRT data to a maximum of r = 0.88 ($p > 0.001$) whilst the model proposed here was able to consistently predict data above r = 0.95, with only one out of six sets of data modelled being below that, at r = 0.86.

The modelling methodology proposed in this thesis gives us the ability to look not only at the ability of a listener to use better-ear listening and binaural unmasking in complex listening situations, but also to use the model's speed to look at a given scenario in much greater detail than would otherwise be practical. This speed may also allow for the model to be applied in such a way that the configuration of sources can be modified interactively with approaching real-time results generated.

7.2 Demonstrated Model Applications

Given the speed and accuracy of the model it has been possible to generate a number of different maps and polar plots to look at the effects of the acoustic space and the

orientation of the listener, or the listener's dining tables, on speech intelligibility, as well as the impact of binaural hearing on cochlear implant users. Using previous models, such as the model of Lavandier & Culling (2010), it was possible in principle to generate these maps, but it would not have been possible to do so at the same speed, which in turn would have limited in practice the number of maps that could be generated, the range of factors that could be examined, and the depth to which these factors could be investigated.

7.2.1 Acoustic Spaces

In anechoic conditions, the benefits afforded to the listener due to better-ear listening and binaural unmasking are not affected by the proximity of the listener to either the target or the interferer, but rather by the azimuthal separation of the sources. This result is reinforced by the polar modelling carried out in chapter 4, where the distance between the listener and both the target and the interferers is kept constant. In this instance, the predicted benefits varied greatly with the orientation of the listener, as well as the relative locations of the interferers. With respect to the number of interferers in the space, the anechoic maps of Chapter 3 showed that the impact of having three interferers spatially distributed within a space rather than one is to decrease the mean predicted TIR_{BE+BU} across the room by 5.8 dB, in the configuration tested (The figure of 5.8 dB is taken from Table A.4, p.189, after correction to compensate for the increase in power due to the increase in the number of interferers, this same correction is applied to all figures quoted below).

Considering first a single interferer, a single reverberant interferer will effectively arrive at the listener from a multitude of angles, and so the ability of the listener to use binaural cues will be impaired. This effect explains in part the difference in the predicted

TIR_{BE+BU} between the anechoic single interferer condition and the reverberant equivalent, 1.2 dB and 0.5 dB respectively. With multiple interferers a similar effect is predicted, with each individual interferer appearing to arrive from many angles, but as their independent TIR_{BE+BU} will already be lower the effect of this is reduced, causing the 2.3 dB reduction rather than the 5.8 dB reduction observed in the anechoic case. In reverberation the addition of extra interferers is less harmful to the listener than in an anechoic environment: for the reverberant situation modelled in chapter 3 the addition of two extra interferers resulted in a reduction of TIR_{BE+BU} of 2.3 dB rather than the 5.8 dB observed in anechoic conditions.

In both anechoic and reverberant conditions it is not simply a matter of multiple interferers being a bad thing, but rather that interferers distributed in azimuth are harmful to intelligibility. Three interferers at differing azimuths to the listener will be of greater detriment to the listener than the same three interferers positioned so they have the same relative azimuth to the listener, at least with noise interferers.

Aside from the overall level of reverberation, the exact acoustics of a room can change the predicted target-to-interferer ratio. Chapter 6 modelled a more realistic configuration of room and sources, looking not only at the differences between an anechoic room and a reverberant room, but also at the effects of where the absorption is placed within a room. The results of the modelling indicate that for a given reverberation time it is better if the absorption is placed on the walls, so that the majority of the reflections occur from the ceiling and the floor.

In the instance of a school classroom, the level of reverberation allowed is specified in Building Bulletin 93 (DfES 2004). Building Bulletin 93 (BB93) also discusses the location of acoustic absorption in a classroom, giving two main approaches; either an

absorptive ceiling or absorptive walls. For the absorptive walls it recommends having the majority of the absorption high up the walls, and focussed on the rear wall. BB93 does not state any preference for either of the two approaches, so it is up to the designer to determine which is the best method to use. Often it may be easier in a classroom situation to install an acoustic ceiling, than to treat the walls. According to the model this approach will reduce the benefits afforded to the listeners through binaural hearing.

7.2.2 Head Orientation

Binaural hearing, or rather the benefit of binaural hearing on the ability of a listener to use optimal head orientations, has been a feature throughout this thesis. Hawley *et al* (2004) measured speech reception thresholds for a range of interferer azimuths and numbers of interferers. These configurations have been modelled and extended upon to investigate the effect of head orientation on speech understanding (Chapter 4). In this modelling 12 polar plots were created, showing the effect of orientation in 12 different configurations of interferer position and number (Figs. 4.2, 4.3, and 4.4). In 8 of the 12 configurations the optimal orientation for the listeners is not facing the target source, but facing in another direction, typically orientating themselves so that the target is to one side, and the interferers to the other.

Whilst it would be possible to empirically measure these results, to do so in any level of detail would be difficult. In the experiments of Hawley *et al* it took one participant between three and six sessions, from 1.5 to 2 hours each, to collect data for 48 conditions. Assuming an average time of 1.75 hours per session, and an average of 4.5 sessions per participant, the average duration to collect 48 conditions will be approximately 7 hours 50 minutes per participant. In the 12 polar plots mentioned above there are a total of 864

conditions, 72 per polar plot. At the speed at which Hawley *et al* were able to collect their data it would take a participant nearly 6 days to test every configuration needed to generate the 12 polar plots, and it would be necessary to test more than one participant. In contrast, the current implementation of the revised model is able to generate all of the 12 maps in approximately 1 hour 20 minutes (dependent on computer hardware).

Previous studies in the literature have focussed on the effect of the spatial separation of sources when the target is located in front of the listener. In comparison, as the model is able to explore the effect of spatial unmasking for any combination of source azimuths, it can show that the conditions that have been empirically tested have not been optimal for showing the potential binaural benefits available to the listener. The only data available currently where the target is not located in front of the listener is that supplied by Juan-Pablo Ramirez of Deutsche Telekom Laboratories (Berlin Institute of Technology), which was used in the polar plot shown in figure 4.1 (p. 84).

The ability to model the head orientation of the listener allows us to look not only at the effect of listener orientation in individual positions, as in each of the polar plots generated in this work, but also to look at listener orientations for many different positions in a space with a fixed set of target and interfering sources. The optimal-listener-orientation room maps generated in Chapter 4 show the effect of listener position on optimal head orientations. For a given configuration of sources, when the listener is close to the target the optimal predicted orientation is to be facing either towards the target, or between target and interferer, but, as the distance increases, the optimum orientation changes so that it is better to have the target to one side of the listener. The reason for this is to do with the relative azimuth separation of the sources. At short distances the relative azimuth separation of the target and interferer(s) is great, so the need to rotate away from

facing the target to obtain benefits of better-ear listening and binaural-unmasking is small. Whereas, at greater distances, the relative azimuth separation of the sources decreases, and the optimum orientation changes. Rather than optimising to use better-ear listening and binaural unmasking it is simply a case of getting the best possible TIR at one ear, either by pointing that ear towards the target, and making use of the directional response of the ear and placing the target in a more sensitive region than the interferer, or by placing the target at exactly 90° to the listener, in the "bright" spot of the occluded ear. Critically, this only holds true in an anechoic space. In a reverberant space the pattern of optimal orientations is less clear, as in Figure 4.11(b), with the proximity of the listener to the walls also affecting the optimal orientation.

Often the orientation of a listener is relatively fixed, either by the design of a room, such as a classroom or lecture theatre, or due to social rules, where it is assumed that a listener will show attention to a talker by facing them, so head orientation is constrained. In these instances, the design of a space will be important, particularly with reference to the positioning of any noise sources such as ventilation systems and projectors. Classically the concern has been for the proximity of the sources to the listener, but as well as proximity we need to consider their relative orientations to the listener, as compared to the target direction.

7.2.3 Cochlear Implants

One offshoot of the initial research into source separations and head orientations was the effects of head orientations for Cochlear Implant users. Given the limitations on speech intelligibility for cochlear implantees even before interfering sources and reverberation are considered, it is all the more important to get the maximum benefits available

out of any given situation.

Where the benefits of bilateral implantation have been assessed, the situations tested have been confined to having the target directly in front of the listener, which is not optimal for normal hearing listeners, let alone cochlear implant users. Our modelling shows that head movements can provide large benefits to the listener. Given that bilateral cochlear implant users report high levels of satisfaction (Noble et al. 2008), which is at odds to the empirical data, it may be that experienced users make use of head orientations, though empirical data needs to be collected in order to substantiate this suggestion.

One issue with the cochlear implant modelling in this thesis is that there is little data available to validate the model against. It is assumed that the listener is able to obtain the same benefit from better-ear listening as a normally hearing listener. That said, the model is capable of accurately predicting the observed better-ear listening benefit in the one configuration of separated target and interferer that has been empirically tested even when using HRTFs from a normal-hearing manikin, which may have different directivity patterns to those found when using the microphones of a cochlear implant.

7.3 Future Developments

The model that has been developed and implemented in this thesis has allowed for the development of a much more detailed picture than would be practical to collect using participant listeners. In particular, the model has generated optimal orientation polar plots, and multiple-interferer optimum-orientation room maps. Whilst the model has proven accuracy and high speed there are a number of areas in which future development might occur, not only in further speed increases, but also in the use of more advanced acoustic modelling software as well as differing ear directionality profiles.

7.3.1 Advanced Acoustic Modelling

Throughout this thesis, all of the impulse responses used in the modelling have been generated using |WAVE (Culling 1996). |WAVE has a number of limitations which may affect the accuracy of the reverberant maps. In particular, |WAVE does not have the ability to use different absorption coefficients at different frequencies, nor different coefficients on different sections of the same surface. Whilst this does not mean that the maps are invalid, given that it is often best to start with a simplified situation in order to see the core effects of the situation at work, it does mean that, in order to look in greater depth at the effects of room architecture, other means of generating the room impulse responses will need to be sought. |WAVE was used for the modelling for a number of reasons, principally due to its ability to batch process large numbers of source / receiver configurations easily. For understanding the core principles behind the effects of reverberation on conversation in rooms, it worked well, but it may be beneficial to replicate the modelling using more sophisticated architectural acoustics modelling software

Current acoustic analysis software suites such as CATT-Acoustic (http://www.catt.se) and Odeon (http://www.odeon.dk) have the ability to generate binaural room impulse responses for a given configuration of sources and receivers. As such, it would be possible to generate the IRs necessary to drive the model and create maps, but with a more complex configuration of reverberation with frequency, as well as more complex architecture. Taking these limitations into account does not invalidate the findings based on the maps in this thesis, which are essentially maps of simple rooms that have equal reverberation times at all frequencies. The effect of reverberation will still occur, though it may vary with frequency, and the same will be true for the effects of having multiple interferers.

7.3.2 Effect of Reverberation on the Target

As described previously, the model currently does not predict the changes to intelligibility due to the direct effect of reverberation on the target speech. Reverberation arriving at the listener shortly after the direct target sound will be useful for understanding it, whilst later arriving sound will tend to temporally "smear" the target sound, reducing its intelligibility. One way in which this might be included in the model is to use the Useful-to-Detrimental ratio.

The useful-to-detrimental ratio considers the energy of sound that arrives within a certain time of the direct sound, typically sound arriving within 80 ms (U_{80}), as useful. This energy reinforces the target. However, energy arriving after that point is considered to be detrimental, and effectively part of the interferer. U_{80} and other measures of the same variety have been shown to be correlated with speech intelligibility scores, across a range of room configurations (Bradley 1986*a*, Bradley 1986*b*, Bradley & Bistafa 2002).

Thus, including the effect of reverberation on the target may be modelled using the principle of the useful-to-detrimental ratio, by considering that any energy arriving after a given point in the target impulse response is part of the interferer complex. Removing this energy from the target and adding it to the interferer impulse response will predict the effect of reverberation on the intelligibility of the target. One extra thing that this would then allow the model to do that it currently cannot is to evaluate the situation where there is no interferer, but rather just a single person speaking in a reverberant space. In terms of architectural acoustics this is often a situation that needs to be tested, such as a teacher giving a lecture in a large hall, where currently the model may predict a higher level of intelligibility than would be observed in practice. There is the need to test this theory using empirical data collected with both targets that are near to the listener, having a high

direct-to-reverberant ratio, and ones that are far from the listener, having a low direct-to-reverberant ratio.

Further to this, the inclusion of the interferer type, such as using real speech rather than speech-shaped noise, along with the effect of reverberation on the target, may change the results of the room modelling. In particular, the predicted TIR at the larger target-listener distances in the reverberant conditions will currently be predicting the "best case" scenario, given that at these locations the direct-to-reverberant ratios, and therefore the useful-to-detrimental ratios, will be low.

7.3.3 Auditory Prostheses and Directional Microphones

Unlike the literature surrounding the effect of spatial separation on speech understanding in normal hearing listeners, there is little available for hearing-impaired listeners, particularly cochlear implant users, and none available that could be used to validate the predictions of the model. Therefore it would be preferable to gain some empirical data to validate the findings, and confirm the benefits of bilateral cochlear implantation over unilateral implantation on the ability of cochlear implantees to use better-ear effects to understand speech when a spatially separated interferer is present.

Aside from the question of whether to use bilateral or unilateral devices it will be worth investigating the use of directional microphones, not only in cochlear implants, but also in acoustic hearing aids. Currently devices such as the Nucleus 5 produced by Cochlear (http://www.cochlear.com) have the ability to change the directional pattern of their microphones in order to help produce the best listening environment for the user. Typically, when the user is trying to listen to a single voice with background interferers the device uses a configuration designed to focus their sensitivity to the front of the listener,

whilst directing a null at the interfering sound. This makes the assumption that sound of value only comes from the front of the listener, which may impact on the modelling carried out in this thesis, particularly the modelling of optimum head orientations for cochlear implant users.

7.3.4 Audio Reproduction

Generally audio reproduction, from watching TV through to live music concerts, is done in a stereo format. Yet often the spatial separation of the sources is limited, and the listening environment less than ideal.

Taking a typical TV as an example, the separation of the speakers in a TV may be in the region of 0.6 m (for a 32" TV), whilst the distance from the viewer to the TV can be large, 3 m for example. In this instance the azimuthal separation of the two speakers is approximately $11.5°$. At best, if we are watching a program where there is a target voice in the presence of an interferer, and each source is panned to one of the two speakers, the benefits of binaural hearing will be minimal. If, instead of using the speakers built into the TV a separate speaker system is used, the benefits that can be afforded the listener can be maximised, without compromising the perception of the visual and the auditory sources being collocated. By modelling a number of possible speaker positions and configurations for a given room it will be possible to decide on the best way to set up the speakers for speech intelligibility.

Aside from the domestic situation above the model may be used not only to predict the intelligibility of a sound reinforcement system, but also to measure it. Hand-held audio test devices, similar to the XL2 produced by NTi Audio (http://www.nti-audio.com/), are currently capable of using specialist test tones in real world situations to measure the

STIPA (Speech Transmission Index for Public Address systems (BSI 2003)). If, rather than test tones, Maximum Length Sequence (MLS) or Pseudeorandom Number (PN) noise was used, impulse responses could be measured (Borish & Angell 1983), which could then be used by the model to calculate the expected TIR. A model that required large amounts of data processing, such as the original model of Lavandier & Culling (2010), would be difficult to implement on a small device, as well as being likely to run slowly on a device with limited processing power.

7.3.5 Further Speed Increases

In this work, the room modelling that has been carried out uses relatively small spaces. If a user of the model wanted to analyse large spaces, and to be able to change things like surface absorption coefficients and see the effects of these changes, the model will need to run quickly. Whilst the current implementation is fast and efficient in comparison to the original model of Lavandier and Culling (2010), there are a number of ways in which it may be made quicker.

7.3.5.1 Working in the Frequency Domain

The current implementation of the model works in both the time domain and the frequency domain. The gammatone filtering of the signal is implemented in the time domain and to calculate the TIR and the B_{BE} the levels of the target and the interferer at each ear are calculated from the gammatone filter output.

At the same time the B_{BU} is calculated using the frequency domain. To calculate the coherence of the interferer in each band, as well as the phase of the target, the interaural cross-correlations of the two signals first need to be calculated. In MatLab this is imple-

mented by using the cross-correlation theorem, which states that the cross-correlation of two signals is equal to the inverse Fourier transform of the product of the forward Fourier transform of one of the signals multiplied by the complex conjugate of the Fourier transform of the second signal. That is:

$$R\{L,R\} = F'\left[(F\{L\})^* \cdot F\{R\}\right] \tag{7.1}$$

where R is the cross-correlation, F and F' the forward and inverse Fourier transforms, L and R the left and right signals, and * represents the complex conjugate.

At the moment, therefore, calculation of a cross correlation involves Fourier transformation of both the left and the right ear channels, and then an inverse Fourier transform. This is carried out for every frequency band and for both the target and the interferer. One way of speeding up this process would be to do a single forwards Fourier transform for each signal prior to filtering into the individual frequency bands, and then doing one inverse Fourier transform for each band to gain the cross-correlation function. By doing this the number of inverse Fourier transforms would stay the same, but the number of forwards Fourier transforms would be reduced to only four, two for the target (left and right) and two for the concatenated interferers (again, left and right), rather than four per frequency band.

Given that using this method of band filtering would occur in the frequency domain it would be possible to also calculate the levels of the target and the interferer also in the frequency domain, removing the need to do any filtering in the time domain.

7.3.5.2 Implementation in C/C++

Whilst MatLab is an efficient medium for developing code there are processing overheads incurred when running a piece of code compared to running an equivalent piece of code directly through the computer operating system. One effect of this is that if the code necessary to run the model were to be re-written in a language such as C or C++ speed increases will be seen, even when running on the same hardware. The additional bonus of writing the code in a language such as C is that with care it can be written to run on any operating system, without needing third-party software.

7.4 Conclusions

The aim of this study has been to produce a prediction methodology capable of quickly and efficiently modelling the benefits of binaural hearing on speech understanding. This model was used to investigate the ways in which reverberation and the configuration of sources in a room affect the ability of a listener to make use of these binaural benefits.

Lavandier and Culling (Lavandier & Culling 2010) proposed a method of predicting the ability of a listener to understand speech in noise, taking into account the effects of better-ear listening and binaural unmasking. They used their model to predict a set of data consisting of an anechoic target and single interferers of differing levels of reverberation. Whilst this model worked well, it was comparatively slow to use. Developing their model further, and modifying it so that the required input data could be minimised without compromising accuracy, has enable more to be done with the model. There are a number of benefits of the revised model over the previous model:

- The model is able to predict a number of empirical data sets from the literature, including sets which include HRTFs rather than being modelled as a pair of microphones in space as was the case of the data of Lavandier and Culling (2010).

- With the increased speed of the revised model it is possible to look at many different configurations of listener, target, and interferers quickly and efficiently, with the ability to generate complex detailed maps of spaces in a short period of time.

With the development of the new model came the ability to look at large numbers of configurations, in a level of detail that, whilst technically possible with the Lavandier & Culling version, would have been highly time consuming and difficult in practice, requiring storage and manipulation of large amounts of data. Because of this, the revised model

is able to tell us things that would otherwise have been impractical to investigate:

- Reverberation reduces the ability of the listener to use better-ear listening and binaural unmasking, particularly at greater distances when the direct-to-reverberant ratio is low.

- In terms of understanding speech in the presence of a noise interferer, reflections from walls are more detrimental than reflections from the ceiling. This goes against the typical thought that it is best to keep reflective walls, in order to maximise benefits arising from early reflections.

- Often it is best to orientate so as not to directly face the target. When close to the target it is best to orientate between the target and the interferers so as to maximise the differences in interaural time and level differences, whilst at greater distances it is better to orientate such that you can maximise the target-to-interferer ratio at one ear given the reduction in better-ear listening and binaural unmasking effects.

- The revised model has the ability to include the effects of cochlear implantation, which had not been explored with the original Lavandier and Culling model. It has been possible to analyse the benefits of bilateral cochlear implantation at a range of listener positions, and many head orientations.

There are a number of possible ways in which future work may build on the present work:

- By collecting empirical data it will be possible to find out if the predicted benefits of bilateral Cochlear Implantation are accurate. If they are, then it might give greater reason to give bilateral cochlear implants rather than unilateral implants.

- Modulated interferers, such as speech, will give very different levels of intelligibility to noise. Using data from the literature and data collected empirically it will hopefully be possible to devise a method capable of including these effects in the model.

- By including the effects of reverberation on the intelligibility of the target. This may be achieved using the Useful-to-Detrimental ratio approach of Bradley (1986*a*, 1986*b*) as discussed previously (Chapter 7.1.2.3).

References

Algazi, V., Duda, R., Thompson, D. & Avendano, C. (2001), The CIPIC HRTF database, *in* 'Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics', IEEE, Mohonk Mountain House, New Paltz, NY, pp. 99–102.

ANSI (1969), *S3.5 - Methods for the Calculation of the Articulation Index*, American National Standards Institue,, New York.

ANSI (1997), *S3.5 - Methods for the Calculation of the Speech Intelligibility Index*, American National Standards Institute, New York.

Badin, P., Tarabalka, Y., Elisei, F. & Bailly, G. (2010), 'Can you 'read' tongue movements? evaluation of the contribution of tongue display to speech understanding', *Speech Communication* **52**(6), 493–503.

Bernstein, L. & Trahiotis, C. (1996), 'On the use of the normalized correlation as an index of interaural envelope correlation', *The Journal of the Acoustical Society of America* **100**(3), 1754–1763.

Beutelmann, R. & Brand, T. (2006), 'Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners', *The Journal of the Acoustical Society of America* **120**(1), 331–342.

Beutelmann, R., Brand, T. & Kollmeier, B. (2010), 'Revision, extension, and evaluation of a binaural speech intelligibility model', *The Journal of the Acoustical Society of America* **127**(4), 2479–2497.

Blauert, J. (1996), *Spatial Hearing*, MIT, Cambridge, MA.

Blauert, J., Brueggen, M., Bronkhorst, A., Drullman, R., Reynaud, G., Pellieux, L., Krebber, W. & Sottek, R. (1998), 'The AUDIS catalog of human HRTFs', *The Journal of the Acoustical Society of America* **103**(5), 3082–3082.

Bond, M., Mealing, S., Anderson, R., Elston, J., Weiner, G., Taylor, R., Hoyle, M., Liu, Z., Price, A. & Stein, K. (2009), 'The effectiveness and cost-effectiveness of cochlear implants for severe to profound deafness in children and adults: A systematic review and economic model.', *Health Technology Assesment* **13**(44).

Boner, C. (1942), 'Performance of broadcast studios designed with convex surfaces of plywood', *Journal of the Acoustical Society of America* **13**, 244–247.

Borish, J. & Angell, J. (1983), 'An efficient algorithm for measuring the impulse response using pseudorandom noise', *Journal of the Audio Engineering Society* **31**(7/8), 478–488.

Bradley, J. (1986*a*), 'Predictors of speech intelligibility in rooms', *The Journal of the Acoustical Society of America* **80**(3), 837–845.

Bradley, J. (1986*b*), 'Speech intelligibility studies in classrooms', *The Journal of the Acoustical Society of America* **80**(3), 846–854.

Bradley, J. & Bistafa, S. (2002), 'Relating speech intelligibility to useful-to-detrimental sound ratios (l)', *The Journal of the Acoustical Society of America* **112**(1), 27–29.

Brokx, J. & Nooteboom, S. (1982), 'Intonation and the perceptual separation of simultaneous voices', *Journal of Phoneics* **10**, 23–36.

Bronkhorst, A. (2000), 'The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions', *Acta Acustica United with Acustica* **86**(1), 117–128.

Bronkhorst, A. & Plomp, R. (1988), 'The effect of head-induced interaural time and level differences on speech intelligibility in noise', *The Journal of the Acoustical Society of America* **83**(4), 1508–1516.

Bronkhorst, A. & Plomp, R. (1992), 'Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing', *The Journal of the Acoustical Society of America* **92**(6), 3132–3139.

BSI (2003), *Bs En 60268-16:2003 - Sound System Equipment: Objective Rating of Speech Intelligibility by Speech Transmission Index*, British Standards Institure.

Buss, E., Pillsbury, H., Buchman, C., Pillsbury, C. & Clark, M. (2008), 'Multicenter u.s bilateral med-el cochlear implantation study: Speech perception over the first year of use', *Ear Hear* **29**, 20–32.

Cherry, E. (1953), 'Some experiments on the recognition of speech, with one and with two ears', *The Journal of the Acoustical Society of America* **25**(5), 975–979.

Culling, J. (1996), 'Signal-processing software for teaching and research in psychoacoustic under unix and x-windows', *Behavior Research Methods, Instruments and Computers* **28**(3), 376–382.

Culling, J., Hawley, M. & Litovsky, R. (2004), 'The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources', *The Journal of the Acoustical Society of America* **116**(2), 1057–1065.

Culling, J., Hawley, M. & Litovsky, R. (2005), 'Erratum: The role head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources', *The Journal of the Acoustical Society of America* **118**(1), 552–552.

Culling, J., Hodder, K. & Toh, C. (2003), 'Effects of reverberation on perceptual segregation of competing voices', *The Journal of the Acoustical Society of America* **114**(5), 2871–2876.

DfES (2004), 'Building bulletin 93 - acousic design of schools'.

Dunn, C., Tyler, R., Witt, S. & Gantz, B. (2006), 'Effects of converting bilateral cochlear implant subjects to a strategy with increased rate and number of channels', *The Annuals of Otology, Rhinology and Laryngology* **115**(6), 425–432.

Durlach, N. (1963), 'Equalization and cancellation theory of binaural masking-level differences', *The Journal of the Acoustical Society of America* **35**(8), 1206–1218.

Durlach, N. (1964), 'Note on binaural masking-level differences at high frequencies', *The Journal of the Acoustical Society of America* **36**(3), 576–581.

Durlach, N. (1972), Binaural signal detection: Equalization and cancellation theory, *in* J. Tobias, ed., 'Foundations of Modern Auditory Theory', Vol. 2, Academic, New York, pp. 371–462.

Edmonds, B. & Culling, J. (2006), 'The spatial unmasking of speech: Evidence for better-ear listening', *The Journal of the Acoustical Society of America* **120**(3), 1539–1545.

Everest, F. (2001), *The Master Handbook of Acoustics*, 4th edn, McGraw Hill, London.

Farina, A. (2000), Simultaneous measurement of impulse respone and distortion with a swept-sine technique, *in* 'Audio Engineering Society Convention 108'.

Fletcher, H. & Galt, R. (1950), 'The perception of speech and its relation to telephony', *The Journal of the Acoustical Society of America* **22**(2), 89–151.

Foxton, J., Riviere, L.-D. & Barone, P. (2010), 'Cross-modal facilitation in speech prosody', *Cognition* **115**(1), 71–78. doi: DOI: 10.1016/j.cognition.2009.11.009.

French, N. & Steinberg, J. (1947), 'Factors governing the intelligibility of speech sounds', *The Journal of the Acoustical Society of America* **19**(1), 90–119.

Gardner, B. & Martin, K. (1994), HRTF measurements of a KEMAR dummy-head microphone, Technical report, MIT Media Lab.

Hawley, M., Litovsky, R. & Culling, J. (2004), 'The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer', *The Journal of the Acoustical Society of America* **115**(2), 833–843.

Hirsh, I. (1948), 'The influence of interaural phase on interaural summation and inhibition', *The Journal of the Acoustical Society of America* **20**(4), 536–544.

Hirsh, I. & Burgeat, M. (1958), 'Binaural effects in remote masking', *The Journal of the Acoustical Society of America* **30**(9), 827–832.

IEEE (1969), 'IEEE recommended practice for speech quality measurements', *IEEE Transactions on Audio and Electroacoustics* **17**, 227–246.

Kidd, G., Mason, C., Brughera, A. & Hartmann, W. (2005), 'The role of reverberation in release from masking due to spatial separation of sources for speech identification', *Acta Acoustica United with Acoustica* **91**, 526–535.

Kollmeier, B. & Wesselkamp, M. (1997), 'Development and evaluation of a german sentence test for objective and subjective speech intelligibility assessment', *The Journal of the Acoustical Society of America* **102**(4), 2412–2421.

Kryter, K. (1962*a*), 'Methods for the calculation and use of the articulation index', *The Journal of the Acoustical Society of America* **34**(11), 1689–1697.

Kryter, K. (1962*b*), 'Validation of the articulation index', *The Journal of the Acoustical Society of America* **34**(11), 1698–1702.

Lavandier, M. & Culling, J. (2010), 'Prediction of binaural speech intelligibility against noise in rooms', *Journal of the Acoustical Society of America* **127**(1), 387–399.

Levitt, H. & Rabiner, L. (1967), 'Predicting binaural gain in intelligibility and release from masking for speech', *The Journal of the Acoustical Society of America* **42**(4), 820–829.

Licklider, J. (1948), 'The influence of interaural phase relations upon the masking of speech by white noise', *The Journal of the Acoustical Society of America* **20**(2), 150–159.

Loizou, P., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J. & Roland, P. (2009), 'Speech recognition by bilateral cochlear implant users in a cocktail-party setting', *Journal of the Acoustical Society of America* **125**, 372–383.

Long, C., Carlyon, R., litovsky, R. & Downs, D. (2006), 'Binaural unmasking with bilateral cochlear implants', *Journal of the Association for Research in Otolaryngology* **7**, 352–360.

Louden, M. (1971), 'Dimension-ratios of rectangular rooms with good distribution of eigentones.', *Acustica* **24**, 101–103.

Lovett, R., Kitterick, P., Hewitt, C. & Summerfield, A. (2010), 'Bilateral or unilateral cochlear implantation for deaf children: An observational study', *Archives of Disease in Childhood* **95**(2), 107–112.

Majdak, J., Laback, B. & Baumgartner, W.-D. (2006), 'Effects of interaural time differences in fine structure and envelope on lateral discrimination in electric hearing.', *Journal of the Acoustical Society of America* **120**, 2190–2201.

Mesgarani, N., Grant, K., Shamma, S. & Duraiswami, R. (2003), Augmented intelligibility in simultaneous multi-talker environments, *in* '2003 International Conference on Auditory Display', International Community for Auditory Display, Boston, MA, pp. 71–74.

Moore, B. & Glasberg, B. (1983), 'Suggested formulae for calculating auditory-filter bandwidths and excitation patterns', *Journal of the Acoustical Society of America* **74**, 750–753.

Muller, J., Schon, F. & Helms, J. (2002), 'Speech understanding in quiet and in noise in bilateral users of the med-el combi 40/401 cochlear implant system', *Ear Hear* **23**, 198–206.

NICE (2009), 'Cochlear implants for children and adults with severe to profound deafness - nice technology appraisal guidance 166', *National Institute of Health and Clinical Excellence* .

Noble, W., Tyler, R., Dunn, C. & Bhullar, N. (2008), 'Unilateral and bilateral cochlear implants and the implant-plus-hearing-aid profile: Comparing self-assessed and measured abilities', *International Journal of Audiology* **47**, 505–514.

Patterson, R., Nimmo-Smith, I., Holdsworth, J. & Rice, P. (1987), An efficient auditory filterbank based on the gammatone function, *in* 'Institute of Acoustics Speech Group on Auditory Modelling', Royal Signal Research Establishment.

Patterson, R., Nimmo-Smith, I., Holdsworth, J. & Rice, P. (1988), 'Spiral vos final report, part a: The auditory filterbank, cambridge electronic design, contract report (apu 2341)'.

Peissig, J. & Kollmeier, B. (1997), 'Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners', *The Journal of the Acoustical Society of America* **101**(3), 1660–1670.

Plomp, R. (1976), 'Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)', *Acoustica* **34**, 200–211.

Plomp, R. & Mimpen, A. (1979), 'Improving the reliability of testing the speech-reception threshold for sentences', *Audiology* **18**, 43–52.

Plomp, R. & Mimpen, A. (1981), 'Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech reception threshold for sentences.', *Acustica* **48**, 325–328.

Psselt, C., Schrter, J., Opitz, M., Divenji, P. & Blauert, J. (1986), 'Generation of binaural signals for research and home entertainment'.

Qin, M. & Oxenham, A. (2003), 'Effects of simulated cochlear-implant processing on speech recognition in fluctuating maskers.', *Journal of the Acoustical Society of America* **114**(446-454).

Rayleigh, L. (1904), *Theory of Sound*, Dover, New York.

Reisberg, D., McLean, J. & Goldfield, A. (1987), Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli., *in* B. Dodd & R. Campbell, eds, 'Hearing by Eye: The Psychology of Lip-Reading', Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 97–113.

Rhebergen, K. & Versfeld, N. (2005), 'A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners', *The Journal of the Acoustical Society of America* **117**(4), 2181–2192.

Schleich, P., Nopp, P. & D'Haese, P. (2004), 'Head shadow, squelch, and summation effects in bilateral users of the medel combi 40/40+ cochlear implant', *Ear Hear* **25**, 197–204.

Sepmeyer, L. (1965), 'Computed frequency and angular distribution of the normal modes of vibration in rectangular rooms.', *Journal of the Acoustical Society of America* **37**(3), 413–423.

Shannon, R., Zeng, F.-G., Kamath, V., Wygonski, J. & Ekelid, M. (1995), 'Speech recognition with primarily temporal cues', *Science* **270**(303-304).

Shinn-Cunningham, B., Ihlefeld, A., Satyavarta & Larson, E. (2005), 'Nottum-up and top-down influences on spatial unmasking', *Acta Acustica United with Acustica* **91**, 967–979.

Steeneken, H. & Houtgast, T. (1980), 'A physical method for measuring speech-transmission quality', *The Journal of the Acoustical Society of America* **67**(1), 318–326.

Tyler, R., Gantz, B., Rubinstein, J., Wilson, B., Parkinson, A., Wolaver, A., Preece, J., Witt, S. & Lowder, M. (2002), 'Three month results with bilateral cochlear implants.', *Ear Hear* **28**, 80S–89S.

van de Par, S. & Kohlrausch, A. (1997), 'A new approach to comparing binaural masking level differences at low and high frequencies', *The Journal of the Acoustical Society of America* **101**(3), 1671–1680.

van Deun, L., van Wieringen, A. & Wouters, J. (2010), 'Spatial speech perception benefits in young children with normal hearing and cochlear implants.', *Ear Hear* **31**(5), 702–713.

van Hoesel, R. (2007), 'Sensitivity to binaural timing in bilateral cochlear implant users', *Journal of the Acoustical Society of America* **121**, 2192–2206.

van Hoesel, R. & Tyler, R. (2003), 'Speech perception, localization, and lateralization with bilateral cochlear implants', *Journal of the Acoustical Society of America* **113**, 1617–1630.

van Wijngaarden, S. & Drullman, R. (2008), 'Binaural intelligibility prediction based on the speech transmission index', *The Journal of the Acoustical Society of America* **123**(6), 4514–4523.

Volkmann, J. (1942), 'Polycylindrical diffusers in room acoustical design.', *Journal of the Acoustical Society of America* **13**, 234–243.

Watkins, A. (2005), 'Perceptual compensation for effects of reverberation in speech identification', *The Journal of the Acoustical Society of America* **118**(1), 249–262.

Watson, C. (2005), 'Some comments on informational masking', *Acta Acustica Univted with Acustica* **91**, 502–512.

Wesselkamp, M. (1994), Messung Und Modellierung Der Verstndlichkeit Von Sprache, PhD thesis.

Wesselkamp, M., Kliem, K. & Kollmeier, B. (1992), Erstellung eines optimierten satztestes in deutscher sprache, *in* B. Kollmeier, ed., 'Moderne Verfahren Der Sprachaudiometrie', Median-Verlag, Heidelberg, pp. 330–343.

Wilson, B., Finley, C., Lawson, D., Wolford, R., Eddington, D. & Rabinowitz, W. (1991), 'Better speech recognition with cochlear implants', *Nature* **352**, 236–238.

Zurek, P. (1993), Binaural advantages and directional effects in speech intelligibility, *in* G. Studebaker & I. Hochberg, eds, 'Acoustical Factors Affecting Hearing Aid Performance', Allyn and Bacon, Needham Heights, MA., pp. 255–276.

Zurek, P. & Durlach, N. (1987), 'Masker-bandwidth dependence in homophasic and antiphasic tone detection', *The Journal of the Acoustical Society of America* **81**(2), 459–464.

Appendix A

Supporting Statistics

A.1 Chapter 2 - The Revised Method

	Interferer		
Condition	Distance (m)	Azimuth (°)	ILD Only?
1	0.65	-25	
2	0.65	0	
3	0.65	25	
4	5	-25	
5	5	0	
6	5	25	
7	0.65	-25	X
8	0.65	0	X
9	0.65	25	X
10	5	-25	X
11	5	0	X
12	5	25	X

Table A.1: Table showing the configuration of interferers for the 12 conditions used in experiment 1 of the unpublished data, (see 2.2.6, 41). Showing interferer distance and azimuth, and whether the IRs were processed to include ILDs only. Target was always located at 0.65m from the listener and an azimuth of 25°

		Target		Interferer	
Condition	Room	Distance (m)	Azimuth (°)	Distance (m)	Azimuth (°)
1	C	0.65	0	1.25	0
2	C	0.65	0	5	0
3	L	0.65	0	2.5	0
4	L	0.65	0	10	0
5	M2	0.65	-25	0.65	25
6	M2	0.65	-25	1.25	25
7	M2	0.65	-25	5	25
8	LH	0.65	-25	0.65	25
9	LH	0.65	-25	2.5	25
10	LH	0.65	-25	10	25
11	LH	0.65	25	0.65	-25
12	LH	0.65	25	5	-25

Table A.2: Table showing the configuration of target and interferer sources for the 12 conditions used in experiment 2 of the unpublished data, (see 2.2.6, 41). Showing target and interferer distance and azimuths.

	Target		Interferer(s)				
Condition	Distance (m)	Azimuth (°)	Distance (m)	Azimuth 1 (°)	Azimuth 2 (°)	Azimuth 3 (°)	ILD only?
1	0.65	25	0.65	-25			
2	0.65	25	0.65	-25	-5		
3	0.65	0	0.65	-25	25		
4	0.65	25	0.65	-25	-5	5	
5	0.65	25	0.65	-25			
6	0.65	25	0.65	-25	-5		
7	0.65	0	0.65	-25	25		
8	0.65	25	0.65	-25	-5	5	
9	0.65	25	0.65	-25			X
10	0.65	25	0.65	-25	-5		X
11	0.65	0	0.65	-25	25		X
12	0.65	25	0.65	-25	-5	5	X
13	0.65	25	0.65	-25			X
14	0.65	25	0.65	-25	-5		X
15	0.65	0	0.65	-25	25		X
16	0.65	25	0.65	-25	-5	5	X

Table A.3: Table showing the configuration of target and interferer sources for the 16 conditions used in experiment 3 of the unpublished data, (see 2.2.6, 41). Showing target and interferer distance and azimuths, as well as indicating whether IRs were processed to contain ILDs only.

A.2 Chapter 3 - Example Room Maps

	Single Interferer		Multiple Interferer	
	Anechoic	Reverberant	Anechoic	Reverberant
Omnidirectional Map	1.2 (7.6)	0.5 (4.9)	-6.4 (6.8)	-5.7 (4.4)
Better-Ear Listening Map	5.4 (7.4)	2.8 (4.6)	-4.2 (6.7)	-4.2 (4.1)
Combined Effects Map	8.1 (7.5)	4.0 (4.4)	-2.5 (6.7)	-3.1 (4.0)

Table A.4: Table showing statistics for the Omnidirectional, Better-Ear, and Combined room maps. Each cell shows the *mean* (*standard deviation*) Target-to-Interferer ratios dB. 4.77 dB of the difference in target-to-interferer ratios between the single and multiple interferer conditions is due to the increased power of the concatenated interferer impulse responses, for direct comparison 4.77dB needs to be added to each of the multiple interferer target-to-interferer ratios.

	Single Interferer		Multiple Interferer	
	Anechoic	Reverberant	Anechoic	Reverberant
Benefit of Better-Ear Listening	0 / 8.8 (4.2 / 2.2)	0.1 / 8.4 (2.3 / 1.1)	0 / 8.2 (2.2 / 2.0)	0 / 8.0 (1.5 / 1.1)
Benefit of Binaural-Unmasking	0 / 3.8 (2.7 / 1.3)	0 / 3.7 (1.2 / 0.6)	0 / 3.7 (1.7 / 1.0)	0 / 3.5 (1.1 / 0.5)

Table A.5: Table showing statistics for the Better-Ear listening and Benefit of Binaural-Unmasking Maps. Each cell shows the *min* / *max* (*mean* / *standard deviation*) of the benefits in dB.

A.3 Chapter 5 - Audiological Implications

	Unilateral Implantaion		Bilateral Implantation	
	Min / Max	Mean / s.d.	Min / Max	Mean / s.d.
Facing Target	-24.1 / 36.1	-0.5 / 6.9	-22.3 / 36.1	3.8 / 6.8
Facing Optimal	-21.3 / 41.3	4.8 / 7.0	-21.0 / 41.3	5.2 / 7.0
Benefit of Optimisation	0 / 12.3	5.3 / 3.8	0 / 10.6	1.4 / 1.8

Table A.6: Table showing the statistics for the unilateral and bilateral CI TIR maps, giving min, max, mean and standard deviations across all points of the maps, figures 5.4 and 5.5

	Min	Max	Mean	s.d.
Facing Target	0	10.5	4.3	3.7
Facing Target	0	1.5	0.4	0.6

Table A.7: Table showing the statistics of the unilateral versus bilateral CI benefit maps, giving min, max, mean and standard deviations across all points of the maps. Figure 5.6